



# Evolution of Mitochondria Reconstructed from the Energy Metabolism of Living Bacteria

Mauro Degli Esposti<sup>1\*</sup>, Bessem Chouaia<sup>2</sup>, Francesco Comandatore<sup>3</sup>, Elena Crotti<sup>2</sup>, Davide Sasseria<sup>3#a</sup>, Patricia Marie-Jeanne Lievens<sup>1#b</sup>, Daniele Daffonchio<sup>2</sup>, Claudio Bandi<sup>3</sup>

**1** Italian Institute of Technology, Genoa, Italy, **2** Department of Food, Environmental and Evolutionary Sciences, University of Milan, Milan, Italy, **3** Dipartimento di Scienze Veterinarie e Sanità Pubblica, University of Milan, Milan, Italy

## Abstract

The ancestors of mitochondria, or proto-mitochondria, played a crucial role in the evolution of eukaryotic cells and derived from symbiotic  $\alpha$ -proteobacteria which merged with other microorganisms - the basis of the widely accepted endosymbiotic theory. However, the identity and relatives of proto-mitochondria remain elusive. Here we show that methylotrophic  $\alpha$ -proteobacteria could be the closest living models for mitochondrial ancestors. We reached this conclusion after reconstructing the possible evolutionary pathways of the bioenergy systems of proto-mitochondria with a genomic survey of extant  $\alpha$ -proteobacteria. Results obtained with complementary molecular and genetic analyses of diverse bioenergetic proteins converge in indicating the pathway stemming from methylotrophic bacteria as the most probable route of mitochondrial evolution. Contrary to other  $\alpha$ -proteobacteria, methylotrophs show transition forms for the bioenergetic systems analysed. Our approach of focusing on these bioenergetic systems overcomes the phylogenetic impasse that has previously complicated the search for mitochondrial ancestors. Moreover, our results provide a new perspective for experimentally re-evolving mitochondria from extant bacteria and in the future produce synthetic mitochondria.

**Citation:** Degli Esposti M, Chouaia B, Comandatore F, Crotti E, Sasseria D, et al. (2014) Evolution of Mitochondria Reconstructed from the Energy Metabolism of Living Bacteria. PLoS ONE 9(5): e96566. doi:10.1371/journal.pone.0096566

**Editor:** Hemachandra Reddy, Oregon Health & Science University, United States of America

**Received:** January 10, 2014; **Accepted:** April 7, 2014; **Published:** May 7, 2014

**Copyright:** © 2014 Degli Esposti et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work has been partially supported by the project BIODESERT GA-245746 "Biotechnology from Desert Microbial Extremophiles for Supporting Agriculture Research Potential in Tunisia and Southern Europe" (European Union) and the Prin 2009 (grant 009L27YC8\_003), from the Italian Ministry of Education, University and Research (MIUR). Work at IIT has been sustained by intramural funds. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: mauro.degliestposti@iit.it

#a Current address: Dipartimento di Biologia e Biotecnologie, University of Pavia, Pavia, Italy

#b Current address: Department of Life and Reproduction Sciences, University of Verona, Verona, Italy

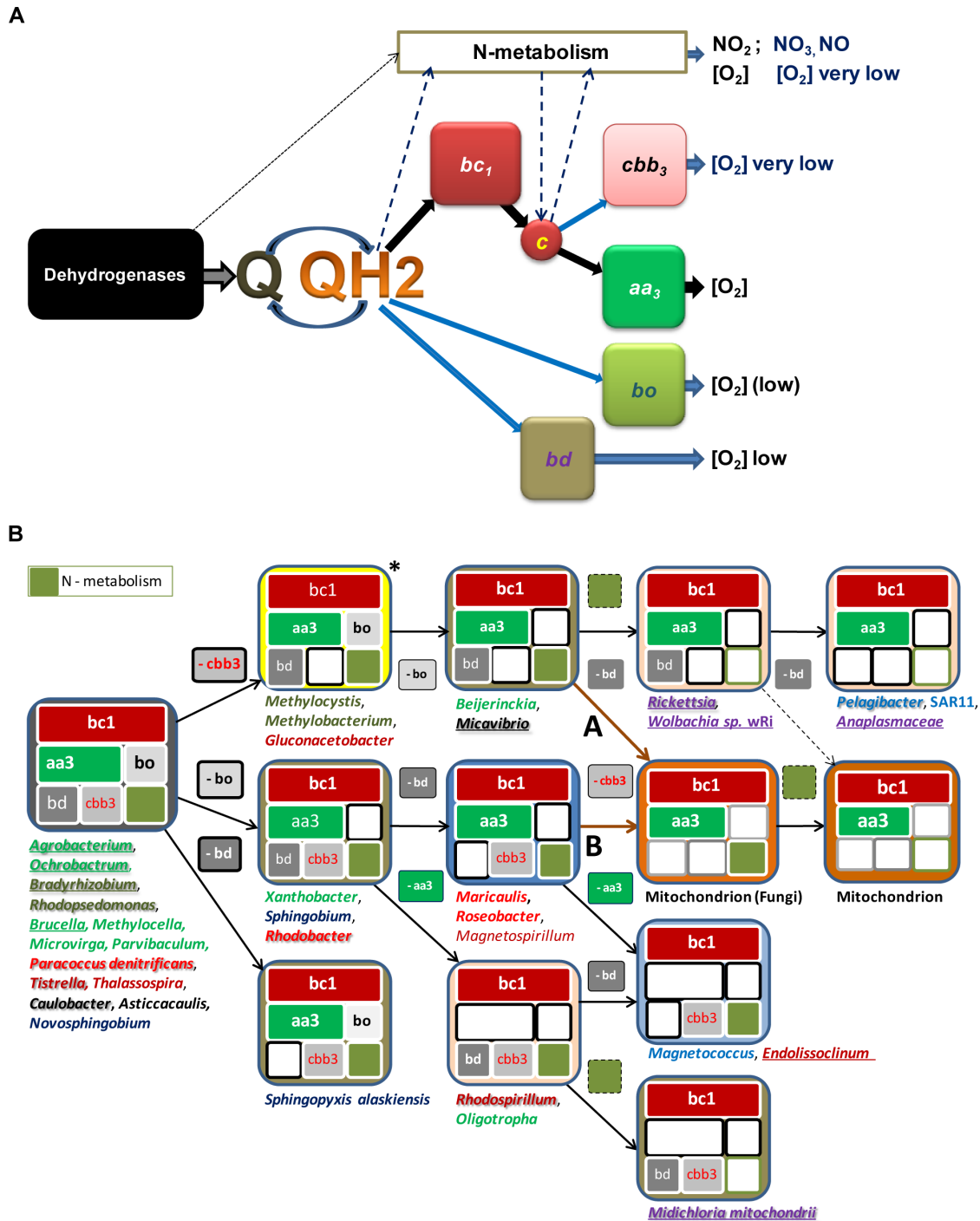
## Introduction

A major concept in biology is that the evolution of eukaryotic cell followed a symbiotic event between diverse microorganisms [1–4]. Mitochondria are the remnants of one of the original partners of this symbiotic event and in all likelihood are related to extant  $\alpha$ -proteobacteria [1–4]. However, the identity of the proto-mitochondrion remains elusive [1]. Phylogenetic studies suggested a relationship with endocellular parasites of the Rickettsiales order [4,5], which has not been confirmed in subsequent reports [6–8]. Indeed, there appears to be a "phylogenetic impasse" in the identification of the partners that merged into the ancestral symbiotic progenitor of current eukaryotic cells [9], partly due to the problem of long branch attraction blurring the true genealogy of living organisms and the fast evolution of mitochondrial DNA [1,10].

The diverse metabolic processes carried out by living bacteria provide complementary approaches to reconstruct key characteristics of the mitochondrial ancestors [11]. Although widely accepted, the reconstruction of proto-mitochondrial metabolism [12] has been partially contradicted by recent evidence suggesting that proto-mitochondria could be related to facultatively anaerobic generalists such as *Rhodobacter* [6–8,10] - which are also capable of

anoxygenic photosynthesis, an autotrophic function that must have been lost early along the evolution of mitochondria. Conversely, this evidence has recently been challenged by controversial reports that aerobic marine organisms such as *Pelagibacter ubique* may be the closest living relatives of mitochondria [13–15]. Other bacterial genera have also been considered to be phylogenetically related, or to display some analogies to the proto-mitochondrion: *Rhodospirillum* on the basis of extensive protein analysis [16]; *Paracoccus* for bioenergy considerations [1], and more recently following the evolution of complex I [17]; *Caulobacter*, on the basis of the sequence similarity of its homologues to the mitochondrial transport protein Tim44 [18]; *Micavibrio*, for its predatory ectoparasite character [19]; the Rhizobiales, *Ochrobactrum* and *Rhodopseudomonas*, for having many proteins in sister position to their mitochondrial homologues [6–8,20]; and finally *Midichloria*, which appears to be the sole representative of the Rickettsiales retaining ancestral features typical of free-living bacteria [21]. The wide diversity of the proposed bacterial ancestors of mitochondria arises from the different approaches of molecular evolution that have been used and the inherent limits of such approaches [1–4].

This work follows a novel approach to identify proto-mitochondrial relatives among extant organisms by focusing on



**Figure 1. Bioenergetic systems of bacteria and mitochondria. A -Terminal respiratory chain of bacteria. 11.** Various bioenergetic systems - membrane redox complexes identified by their common name and different colours - carry out the oxidation of quinols (QH<sub>2</sub>) reduced by dehydrogenases. Besides oxygen (O<sub>2</sub>), nitrogen compounds can function as electron acceptors for the oxidation of dehydrogenases (dotted arrow), quinols and cytochrome c (dashed dark blue arrows), in reactions catalysed by enzyme complexes such as *Nrf* nitrite reductase [32], which are included within the N-metabolism system. Thick black arrows indicate electron transport in aerobic bacteria and mitochondria. Blue arrows indicate other electron transport pathways of facultatively anaerobic bacteria. **B - Pathways of mitochondrial bioenergetic evolution.** The bioenergetic systems illustrated in A are indicated by the coloured modules (with size proportional to their bioenergetic output) within the boxes representing the bioenergetic subset of each organism or organelle. Mitochondria of fungi and heterokont microorganisms differ from those of other eukaryotes for the presence of elements of N-metabolism. Representative taxa with fully sequenced genome are listed beneath each subset. The pathways of mitochondrial evolution are deduced by connecting these subsets with stepwise loss of a single bioenergetic system. Microorganisms underlined are symbionts or pathogens. Bacteria in embossed typeface have been proposed as ancestors or relatives of mitochondria (see Table S1 in File S1 for specific references). Dark brown arrows A and B indicate the pathways leading to fungal mitochondria. The pathway between the Rickettsia subset and that of mitochondria (dashed arrow) can be discounted, since the symbiotic event occurred only once [1,5,6,10,48]. \* indicates the subset from which other pathways depart (Figure S1 in File S1). doi:10.1371/journal.pone.0096566.g001

**Table 1.** Elements of N-metabolism that are shared by bacteria and eukaryotes.

Taxonomic group and organism	NAD(P)H dependent, assimilatory			PQQ-dehydrogenase
	<i>NirB</i>	<i>NirBD</i>	<i>NiaD</i> -related proteins	<i>MxaF</i>
<b>methanotrophs &amp; methylotrophs</b>				
<i>Methylocystis</i> sp. SC2	yes		1 domain	yes
<i>Methylocystis parvus</i>			precursor & 1 domain	yes
<i>Methylosinus trichosporium</i> OB3b	yes		1 domain	yes
<i>Methylosinus</i> sp. LW4			1 domain	yes
<i>Methylocella silvestris</i> BL2	yes		1 domain	yes
<i>Beijerinckia indica</i> *		yes	precursor & 2 domains	yes
<i>Microvirga</i> sp. WSM3557	yes			yes
<i>Methylobacterium extorquens</i> DM4			3 domains	yes
<i>Methylobacterium extorquens</i> PA1			3 domains	yes
<i>Methylobacterium extorquens</i> AM1			2 domains	yes
<i>Methylobacterium extorquens</i> CM4				yes
<i>Methylobacterium extorquens</i> DSM 13060				yes
<i>Methylobacterium nodulans</i> ORS 2060			2 domains	yes
<i>Methylobacterium populi</i> BJ001			2 domains	yes
<i>Methylobacterium radiotolerans</i> JCM 2831			2 domains	yes
<i>Methylobacterium mesophilicum</i> SR1.6/6			2 domains	yes
<i>Methylobacterium</i> sp. GXF4			2 domains	yes
<i>Methylobacterium</i> sp. 88A			2 domains	yes
<i>Methylobacterium</i> sp. 4–46				yes
<i>Xanthobacter autotrophicus</i> Py3	yes			yes
<i>Hyphomicrobium denitrificans</i> 1NES1	yes			yes
<b>Bradyrhizobiaceae</b>				
<i>Nitrobacter winogradskyi</i> Nb-255	yes			
<i>Nitrobacter hamburgensis</i> X14	yes			
<i>Nitrobacter hamburgensis</i> sp. Nb-255	yes			
<i>Oligotropha carboxidovorans</i> OM4 & OM5	yes			
<i>Rhodopseudomonas palustris</i> BisA53			2 domains	yes
<i>Rhodopseudomonas palustris</i> BisB18			1 domain	yes
<i>Rhodopseudomonas palustris</i> TIE-1			2 domains	
other 4 <i>Rhodopseudomonas palustris</i>			1 domain	
<b>Rhodospirillales</b>				
<i>Granulibacter bethesdensis</i> CGDNIH1	yes		2 domains	yes
<i>Commensalibacter intestini</i> A911	yes			
<i>Acidocella</i> sp. MX-AZ02	yes		1 domain	
<i>Acidiphilium multivorum</i> AIU301	yes			
<i>Acidiphilium cryptum</i> & sp. PM	yes		1 domain	
<i>Gluconobacter oxydans</i> H24		yes	precursor & 2 domains	
<i>Gluconobacter frateurii</i> NBRC 103465		yes	precursor	
<i>Gluconacetobacter oboediens</i> 174Bp2		yes	precursor & 2 domains	
<i>Acetobacter pasteurianus</i> IFO 3283-01/32		yes	precursor	
<i>Acetobacter aceti</i>		yes	precursor & 1 domains	
<i>Gluconacetobacter europaeus</i> LMG 18494		yes	precursor	
<i>Gluconacetobacter diazotrophicus</i> PA15			2 domains	
<i>Acetobacter pomorum</i> DM001		yes		
<i>Acetobacter tropicalis</i> NBRC 101654		yes		
<i>Asaia platicody</i>		yes	precursor	
<i>Saccharibacter</i> sp.	yes		2 domains	

Table 1. Cont.

Taxonomic group and organism	NAD(P)H dependent, assimilatory			PQQ-dehydrogenase
	<i>NirB</i>	<i>NirBD</i>	<i>NiaD</i> -related proteins	<i>MxaF</i>
<i>Tistrella mobilis</i> KA081020–065	yes		2 domains	
<i>Azospirillum lipoferum</i> 4B	yes		1 domain	yes
<i>Azospirillum amazonense</i> Y2	yes			
<i>Azospirillum brasilense</i> Sp245	yes			
<i>Azospirillum</i> sp. B510	yes			
<i>Caenispirillum salinarum</i> AK4	yes			
<i>Thalassospira profundimaris</i> WP0211	yes			
<i>Thalassospira xiamenensis</i> M-5	yes			
<i>Magnetospirillum magneticum</i> AMB-1	yes			
<i>Magnetospirillum</i> sp. SO-1	yes			
<i>Magnetospirillum gryphiswaldense</i> MSR-1	yes			
<b>Rhodobacterales</b>				
<i>Oceanicola granulosis</i>			1 domain	
<i>Oceanicola</i> sp. S124	yes			
<i>Octadecabacter antarcticus</i> 307	yes			
<i>Paracoccus denitrificans</i> PD1222	yes			
<i>Roseobacter denitrificans</i> OCh114	yes			
<i>Roseobacter litoralis</i> Och 149	yes			
<i>Jannaschia</i> sp. CCS1	yes			
<b>Rhizobiales (other)</b>				
<i>Marteella mediterranea</i>			precursor	
<i>Aureimonas ureilytica</i>			2 domains	
<i>Sinorhizobium meliloti</i> 1021	yes		2 domains	
<i>Rhizobium leguminosarum</i> bv. <i>trifolii</i> WSM1325	yes			
other 32 Rhizobiales	yes			
<b>Sphingomonadales &amp; Caulobacterales</b>				
<i>Novosphingobium nitrogenifigens</i>			precursor	
<i>Sphingomonas</i> sp. 17			2 domains	
<i>Sphingomonas</i> sp. PAMC26621			1 domain	
<i>Sphingopyxis alaskensis</i> RB2256	yes			
other 19 Sphingomonadales & 6 Caulobacterales	yes			
<b>total <math>\alpha</math>-proteobacteria</b>	<b>ca. 100</b>	<b>10</b>	<b>12 precursors</b>	
<b>Eukaryotes</b>				
<i>Aspergillus fumigatus</i>		yes	yes	
other 130 fungi (predominantly Ascomycetes)		yes	yes	
<i>Ectocarpus silicosus</i>		yes	yes	
plus other 8heterokonts	(1 yes)	yes	yes	
<i>Aureococcus anophagefferens</i>		yes	yes & 2 domains	
<i>Acanthamoeba castellanii</i>			yes	
<b>total Eukaryotes</b>	<b>1</b>	<b>140</b>	<b>141</b>	

Proteins closely related to *NirB*, *NirBD*, *NiaD* and *MxaF* are annotated as **yes**, or **precursor** in the case of *Nas/CysJ* nitrate reductase (Fig. 2). The column of *NiaD*-related proteins also lists the number of *NiaD* **domains** that have homologues proteins in each organism, e.g. flavohaem (cf. Fig. 2C).

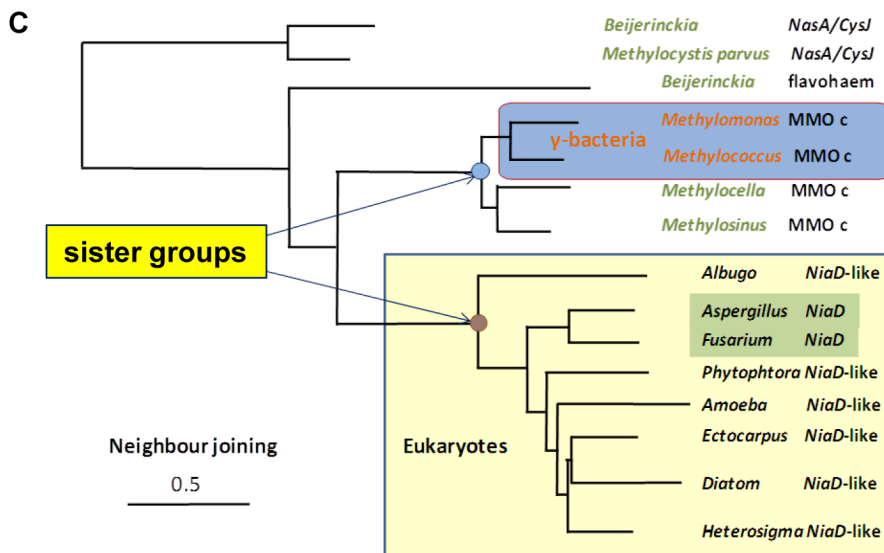
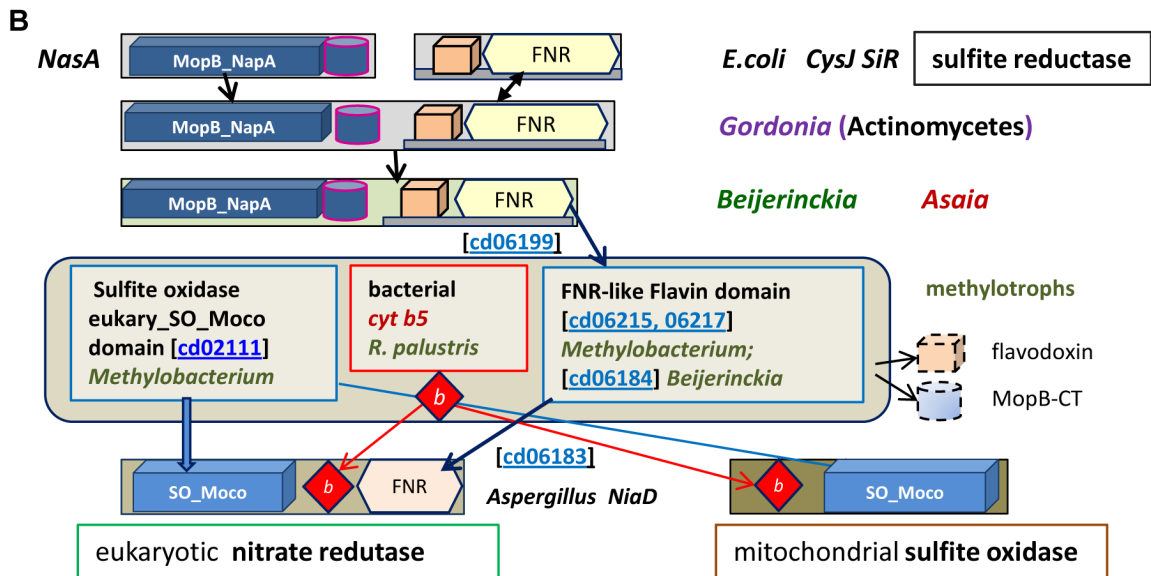
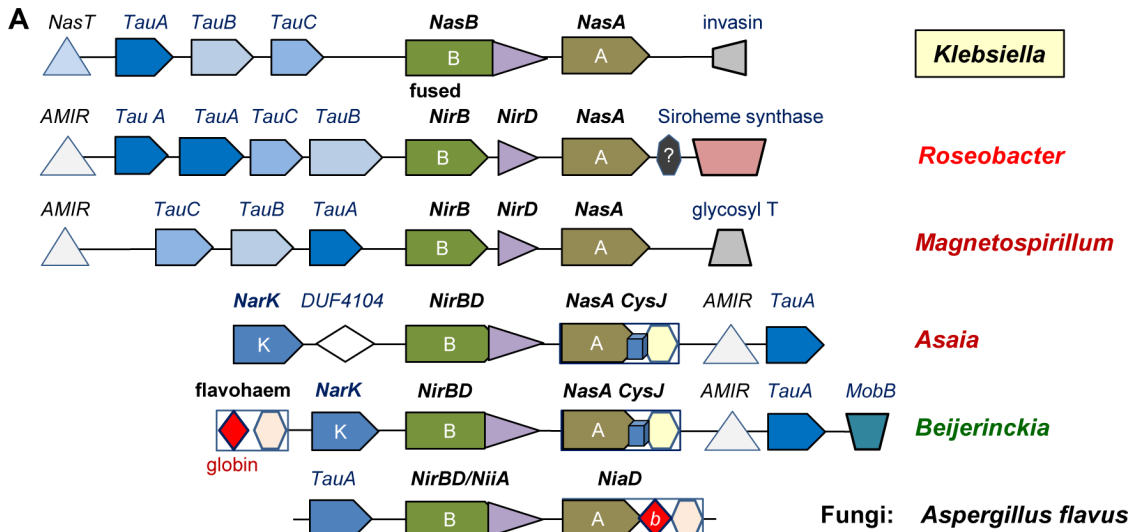
\*Its close relative *Beijerinckia mobilis* has been reported to grow on methanol and possess *MxaF*.

doi:10.1371/journal.pone.0096566.t001

the bioenergetic systems that are common between mitochondria and bacteria. An enormous increase in bioenergy production constitutes the major advantage gained in the endosymbiotic event that led to the evolution of eukaryotic cells [2]. Consequently, the mitochondrial systems that generate most cellular bioenergy must

define the minimal bioenergetic capacity of proto-mitochondria. Whereas aerobic  $\alpha$ -proteobacteria such as *Pelagibacter* present the same two bioenergetic systems of animal mitochondria [4,12], other proposed ancestors of mitochondria such as *Rhodospseudomonas palustris* [6–8] possess four additional bioenergetic systems in their





**Figure 2. Graphical representation of assimilatory nitrate reduction in protists and  $\alpha$ -proteobacteria. A – The diagram shows the gene clusters of assimilatory, NAD(P)H-dependent nitrate reduction in bacteria and eukaryotes. The various elements of *Nas* operon of *Klebsiella* [36] and the *NiIA-NiAD* operon in fungi [35] are colour coded as indicated in the quadrant on the top right. B – Possible molecular evolution of fungal *NiAD* nitrate reductase. Each domain is identified by a specific symbol - see the text for details. C – Representative distance tree of various proteins containing the bacterial FNR-like conserved domain. The tree was obtained with Neighbour Joining (maximal distance 0.9) using the DELTAST program [80] with methane monooxygenase subunit c of *Methylorella silvestris* (MMOC, Accession: YP\_002361598) as query. This reductase subunit of methane monooxygenase contains a FNR-like domain similar to that of assimilatory nitrate reductases [43] lying in a sister group as indicated. doi:10.1371/journal.pone.0096566.g002**

terminal respiratory chain (Fig. 1A). These systems are characteristic of bacteria living under anaerobic or micro-oxic conditions, exploiting also bioenergy-producing elements of N-metabolism which are partially retained in some eukaryotic microorganisms [10,22,23]. It is thus likely that the current bioenergetic portfolio of mitochondria has evolved from a larger genomic endowment of bioenergetic systems which has been reduced via sequential loss.

We have reconstructed the possible pathways of this sequential loss leading to the bioenergetic systems of current mitochondria by evaluating all the genomes of  $\alpha$ -proteobacteria which are currently available. Results obtained with complementary approaches then converged in indicating that methylotrophic  $\alpha$ -proteobacteria could be the closest living relatives to proto-mitochondria, while excluding the majority of bacteria previously proposed as mitochondrial relatives.

## Results and Discussion

### 1.1 Reconstructed pathways of bioenergetic evolution of bacteria into mitochondria

The bioenergetic capacity of mitochondria has been instrumental in the evolution of eukaryotic cells and complex life forms [1–3]. It is generally assumed that proto-mitochondria had an aerobic energy metabolism equivalent to that of today's mitochondria [1,4,12], with the central part of the respiratory chain consisting of ubiquinol-cytochrome *c* reductase (the cytochrome *bc*<sub>1</sub> complex) and a single terminal oxidase, cytochrome *aa*<sub>3</sub> oxidase (Fig. 1A). However, geophysical evidence indicates that proterozoic oceans were essentially anoxic during the period in which the eukaryotic cell evolved [24]. Consequently, it is likely that proto-mitochondria were adapted to different levels of environmental oxygen, exploiting also the terminal oxidases of facultatively anaerobic bacteria to obtain bioenergy [10]. For example, *Rhodospirillum rubrum* strains possess cytochrome *bd* and *bo* ubiquinol oxidases [25,26], plus an additional cytochrome *c* oxidase of the *cbb*<sub>3</sub> type [27] (Fig. 1B). Endocellular parasites have the *bd* ubiquinol oxidase either alone (in several species of *Rickettsia* [28]) or together with *cbb*<sub>3</sub> oxidase (in *Midichloria mitochondrii* [21]). Other organisms, moreover, possess proteins of the anaerobic bioenergetic process of denitrification, which are found also in mitochondria of fungi that can adapt to anaerobiosis [10,23,29].

Fungi and heterokont protists additionally possess an assimilatory nitrite reductase which is involved in ammonia fermentation, *NirB* fused with *NirD* [23,29] – hereby defined as *NirBD*. In some bacteria, this NAD(P)H-dependent enzyme forms part of the nitrogen cycle that enables their growth from the oxidation of methane or ammonia, the oxidation of C1 compounds such as methanol (methylotrophy) and ammonification of nitrite [30–32]. Because various elements of this nitrogen cycle are associated with bioenergy production [23,29–32], we have considered them within the broad bioenergetic system of N-metabolism (Fig. 1).

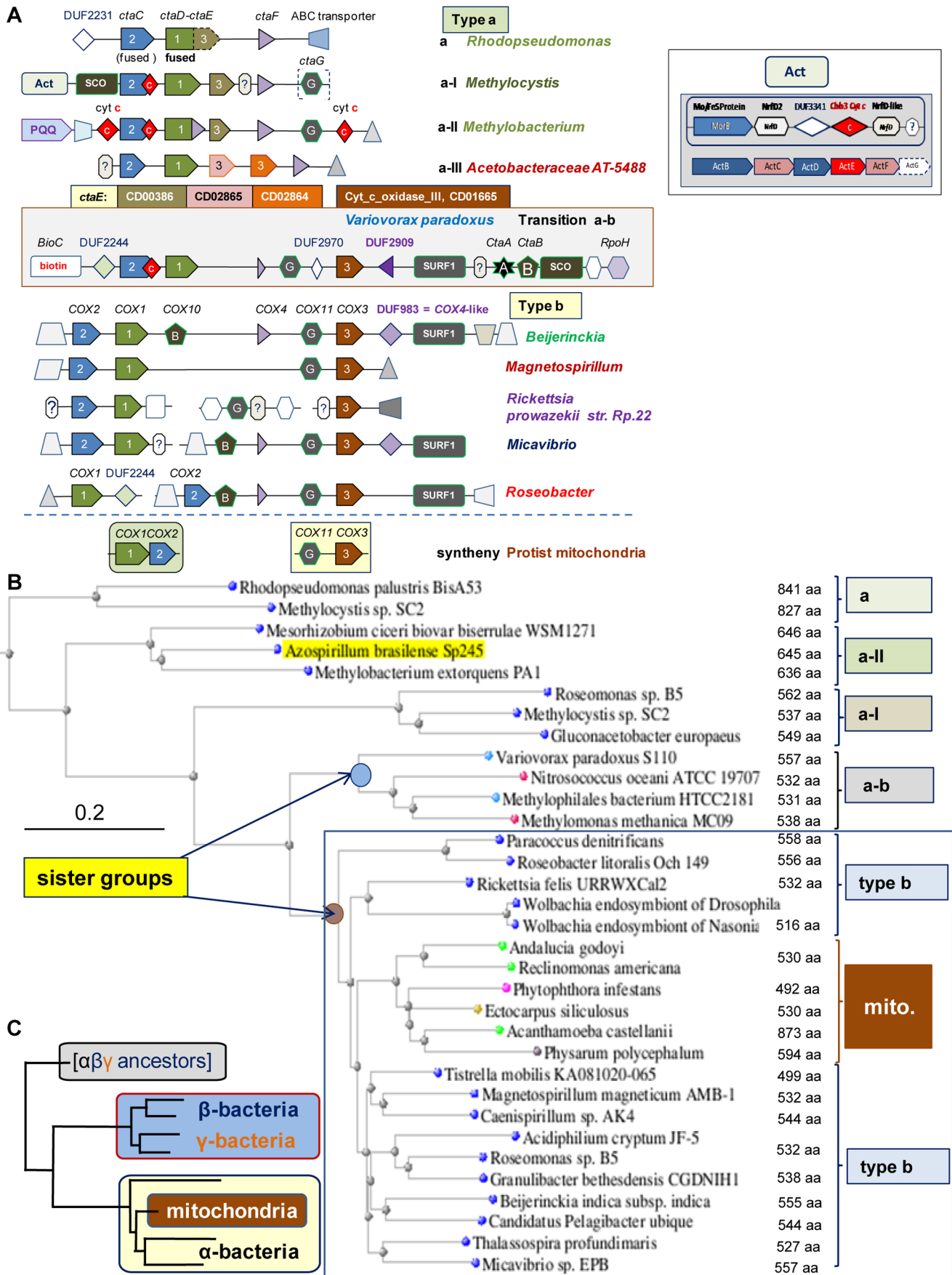
The metabolic versatility of current bacteria suggests that the ancestors of  $\alpha$ -proteobacteria had six bioenergetic systems from ubiquinol to oxygen (Fig. 1B), like diverse extant bacteria (Table S1 in File S1). To deduce the pathways of differential loss

that led to the reduced subset of current mitochondria, we have developed a model based upon the bioenergetic systems coded in all available genomes of  $\alpha$ -proteobacteria, including those we have recently sequenced (*Asaia platycody* and *Saccharibacter sp.* [22]). For parsimony, we allowed only single-step connections between the various subsets, thus obtaining two alternative pathways which directly lead to the subset of bioenergetic systems that is present in contemporary mitochondria of fungi and protists (Fig. 1B, cf. Fig. S1 in File S1). Pathway A stems from the subset present in predatory *Micavibrio* [19] and also *Beijerinckia indica*, a metabolically versatile organism closely related to methylotrophs [33] which has been shown to possess several proteins strongly related to their mitochondrial homologues [8]. Alternative pathway B originates from the subset present in some *Magnetospirillum* species and two Rhodobacterales (Fig. 1B): *Roseobacter litoralis*, which retains a functional photosynthetic apparatus, and *Maricaulis maris*, which has a dimorphic biological cycle. The loss of N-metabolism from the *Micavibrio/Beijerinckia* subset leads to the subset of *Rickettsia* [28] and *Wolbachia* organisms which retain the *bd* ubiquinol oxidase system (Fig. 1B). The loss of this bioenergetic system would also lead to the subset of metazoan (but not fungal) mitochondria, a possibility considered unlikely in view of the unique symbiotic event producing mitochondria [1,2,10]. Moreover, it occurs in related species of the same Rickettsiales order (Fig. 1B) and other taxa, for example within the *Bartonella* genus (Fig. S1 in File S1), suggesting phenomena of convergent evolution.

### 1.2 Testing the alternative pathways for mitochondrial bioenergy evolution

So, comparative genomic analysis has allowed a reconstruction of two possible reductive pathways in the bioenergetic capacity of bacteria evolving into mitochondria (Fig. 1). How can we establish which of these pathways is most likely, and thus identify extant models for proto-mitochondria? Probabilistic approaches based upon the frequency of gene loss from each subset would not produce conclusive evidence, because of the biased phylogenetic distribution of available bacterial genomes. We have then carried out the classical phylogenomic approach of computing the overall relationships of the organisms in the model of Fig. 1B by using concatenated proteins that are common to most eubacteria (cf. Ref. [21]). Although the obtained trees could be globally consistent with the sequence of either pathway A or B, they did not offer discriminatory evidence in favour of one or the other, while consistently placing *Midichloria* and other Rickettsiales close to the mitochondrial clade. This tree topology has been reported before [1,4,5,21] but is inconsistent with our new model of Fig. 1B and other evidence [1], as discussed above.

We next followed the alternative approach of exploiting the molecular diversity of key bioenergetic proteins, including their multiple duplication [34]. To enhance the discriminatory power of this approach, we have chosen proteins of energy metabolism that have a clear bacterial origin, but are encoded or located in different compartments of eukaryotic cells (cf. [34]). The hypothesis underlying our approach is that such diverse proteins,



**Figure 3.  $\alpha$ -proteobacteria have different types of COX operons and catalytic subunits of  $aa_3$  oxidase. A - Graphical representation of  $aa_3$  oxidase gene clusters.** The different COX clusters of  $\alpha$ -proteobacteria are classified by considering gene sequence variations and the features of flanking genes (see also “Classification of bacterial COX operons” in File S1). Specific graphical symbols identify COX subunits as indicated; other types of proteins are labelled as follows: white hexagon, enzyme working with RNA or DNA; red diamond with enclosed c, cytochrome c type protein; truncated triangle pointing left, ABC transporter/permease; grey sharp triangle, transcription regulator; PQQ, PQQ-dependent dehydrogenase; white diamond, protein belonging to a DUF family [41], e.g. DUF983; question mark within hexagon, completely unknown protein. Note that *SURF1* (Surfeit locus protein 1) and *SCO* (Synthesis of cytochrome c oxidase) are also involved in the biogenesis of oxidases. Distance between genes is arbitrary. COX operon **type a-I** is attached to a *Nrf*-like gene cluster, also called Alternative Complex III or Act [50], containing two homologues of the membrane subunit *NrfD* (called *NrfD2* and *NrfD*-like here, as shown at the side of the figure). The synthetic diads of protist mitochondria [48] are shown below the blue line. Each of the recognised subfamilies of COX3 [41] is represented by a different colour, as indicated in the middle of the illustration. **B - Representative distance tree of COX1 proteins.** The tree was obtained with Neighbour Joining (maximal distance 0.9) using the DELTAST program [80] with the COX1 protein of *Methylobacterium extorquens* PA1 (Accession: YP\_001637594) as query. The group containing bacterial and mitochondrial proteins (mito.) is enclosed in the blue square. Protein length and type of COX operon are annotated on the right of the tree. **C - Simplified pattern of typical phylogenetic trees of COX1 proteins.** The tree is modelled to match distance trees of nitrate reductase (Fig. 2C) and COX1 (part B). Branch length is arbitrary. doi:10.1371/journal.pone.0096566.g003

as well as their genetic clusters, would present transition forms between bacteria and mitochondria predominantly in those organisms that are close to the proto-mitochondrial lineage.

## 2. Molecular evolution of assimilatory N metabolism

The first bioenergetic system we considered is N metabolism, the presence or absence of which sharply determines the pathways leading to the mitochondria of fungi and metazoans (Fig. 1B). As mentioned above, fungi and heterokonts possess the assimilatory, NAD(P)H-dependent nitrite reductase *NirBD* [35], a cytosolic enzyme which is common among facultatively anaerobic  $\gamma$ -proteobacteria such as *Klebsiella*, where it was originally called *NasB* [36]. Structurally, *NirBD* is characterised by the fusion of the small protein *NirD* - belonging to the Rieske superfamily of Fe-S proteins coordinated by histidines and cysteines [37] - at the C-terminus of the *NirB* protein, which catalyses the reduction of nitrite and is structurally related to sulfite reductase (*Sir*) [38]. Interestingly, the distribution of *NirB* is restricted to a relatively narrow group of facultatively anaerobic bacteria [38,39], but that of *NirBD* is much narrower (Table 1). After finding *NirBD* in the genome of *Asaia*, we detected only ten homologous genes among  $\alpha$ -proteobacteria - compared with over one hundred in fungi (Table 1), all arranged in similar gene clusters comprising a regulator, nitrate transporters and an assimilatory nitrate reductase. The gene clusters are related to the *Nas* operon of *Klebsiella* (Fig. 2A), with its most compact version being present in fungi and Oomycetes [35].

Among the bacteria associated with pathway A and B in Fig. 1B, only *Beijerinckia* possesses *NirBD* and its cognate gene cluster. *Roseobacter litoralis* and *Magnetospirillum* have *NirB* within an operon similar to that of *Klebsiella* (Fig. 2A), whereas *Maricaulis* and *Micavibrio* do not have the same genes. This situation may well arise from secondary loss of metabolic traits in ecologically specialised organisms such as dimorphic *Maricaulis* and predatory *Micavibrio*. To gain further phylogenetic information, we then exploited the rare occurrence of *NirBD* and its associated nitrate reductase among  $\alpha$ -proteobacteria (Table 1), evaluating the molecular evolution of these modular proteins. The structure of *NirBD* is conserved in  $\alpha$ -proteobacteria and eukaryotes [35] and apparently derives from *NirB* precursors that are present in methylotrophs such as *Methylocystis* (Fig. 2, cf. [35]).

Conversely, the structure of the large protein functioning as nitrate reductase in the *NirBD* gene cluster of  $\alpha$ -proteobacteria resembles that of nitrate reductases from ancient bacteria such as *Gordonia*, which contains three redox modules formed by distinct domains. A typical Molybdenum cofactor-binding domain (Moco) occupies the N-terminus and includes a terminal part binding another molybdopterin cofactor as in *NapA* (periplasmic) and *NasA* (cytoplasmic) reductases [36–40]. This is followed by an intermediate domain homologous to the small redox protein flavodoxin

(Fig. 2B top, cf. [38]). The C-terminus then contains a flavoprotein reacting with the electron donor NAD(P)H which, in combination with flavodoxin, forms a domain closely related to sulfite reductase *CysJ* of *E.coli* (represented by a grey bar in Fig. 2B, cf. [38]). The *CysJ*-related domain belongs to the superfamily of Ferredoxin Reductase-like domains, cd 00322 FNR-like [41], which includes also the C-terminal domain of fungal nitrate reductase, *NiaD* [35,40].

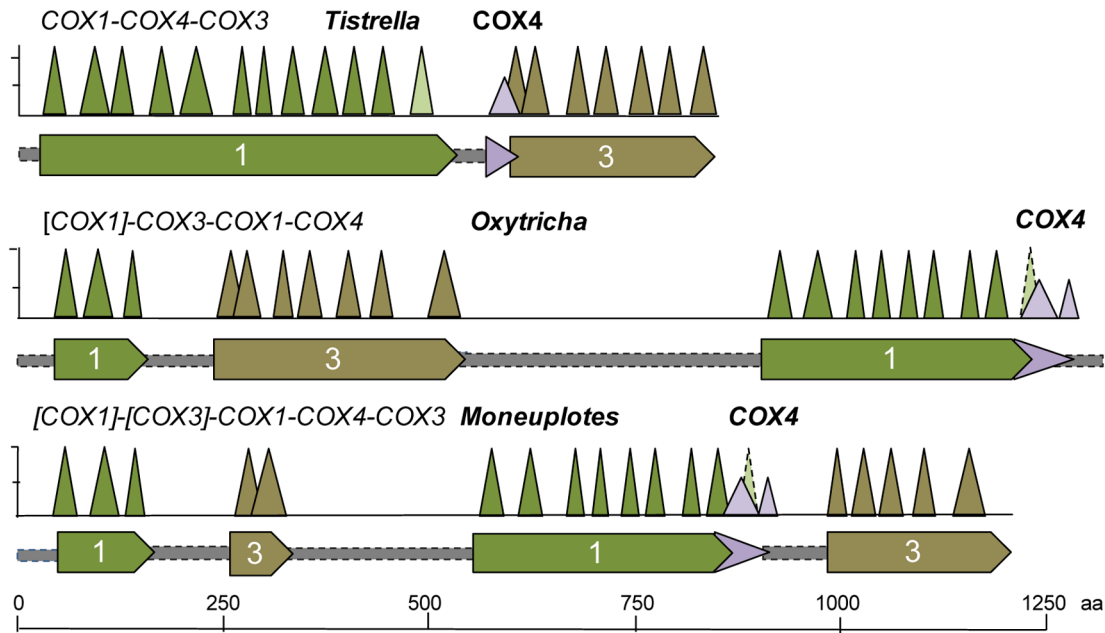
Although the fine structure of the FNR-like domain indicates two separate subfamilies, cd01699 *Sir*\_like for the *NasA/CysJ* bacterial proteins and cd06183 *cytb5\_reductase\_like* for the eukaryotic proteins, our detailed sequence comparison uncovered phylogenetic relationships with other bacterial proteins belonging to the same superfamily. In particular, flavodoxin reductases of the genus *Methylobacterium* and the reductase subunits of soluble methane monooxygenase [42,43] (MMO, present also in close relatives of *Beijerinckia* such as *Methylocella*) were consistently found in sister clades to *NiaD* and related proteins of fungi, heterokonts and *Acanthamoeba* (Fig. 2C and Table 1). Moreover, the flavohaem oxidoreductase of *Beijerinckia* (accession YP\_001833084), which contains a cytochrome *b*-related globin followed by a FNR-like domain, was found in an intermediate position between the *NiaD*-containing clade and the *NasA-CysJ* reductases of *Beijerinckia* and *Methylocystis parvus* (Fig. 2C). Notably, the gene of this protein is located at the beginning of *Beijerinckia* nitrate assimilation operon (Fig. 2A). Its Nitric Oxide dioxygenase activity is also similar to that of the hybrid nitrate reductase of microalgae from the heterokont group, e.g. *Chattonella subsalsa* (protein NR2-2/2HBnN, accession: AER70127), which possess both a cytochrome *b*<sub>5</sub> and a globin in the intermediate domain [44]. These flavoproteins, therefore, could be considered transition forms between *NapA/CisJ* reductases and eukaryotic assimilatory nitrate reductases.

In further support of the modular similarity between bacterial and eukaryotic NAD(P)H-dependent nitrate reductases, we have found that the Moco domain of *NiaD*-like eukaryotic proteins is present also in the sulfite oxidase of methylotrophs such as *Methylobacterium mesophilicum* and *extorquens* (accession: WP\_010685750 and WP\_003602739, respectively - Table 1 and Fig. 2B). Moreover, the genome of *Methylobacterium extorquens PA1* encodes a protein that is partially similar to bacterial cytochrome *b*<sub>5</sub> (accession: YP\_001638730), which is present only in *Rhodospseudomonas palustris* among  $\alpha$ -proteobacteria (Fig. 2B and data not shown). Consequently, all three functional domains of eukaryotic assimilatory reductases have homologous proteins in extant  $\alpha$ -proteobacteria, particularly among those with methylotrophic metabolism, as indicated by the presence of the signature methanol dehydrogenase *MxaF* [45] (Table 1). Hence, our data suggests that *NasA-CisJ* reductases of *Beijerinckia* and acetic acid bacteria, e.g. *Asaia*, represent the likely **precursors** of eukaryotic,

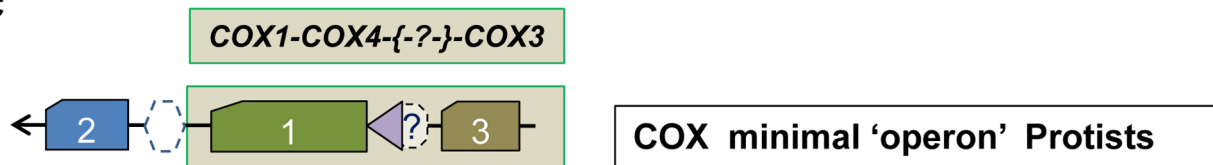
A

Beef mitos	COX3	8	TM1-----TM2-----	negative side
Paracoccus	COX3	13	PSI--WPLFGAIGAFVMLTGAVA--WMKGITFFGLPVEGFWMLIGLVGVLYVMFCWVADVVEGET----GEHT--	
Andalucia	COX3	15	PSP--WPLFAAASVFSLVIGGVM--YM-----HSYENAN--LVFSSGLISLLYVMFVWRDIVERGT--YQGTYYQGHHTVE	
Physarum	COX3	11	PSP--WPIVSAAALLGLTIGGV--SY-----FHSFNNGIYLLTSSFILLAILAGF--WRDLIREGTY----LHNHTKE	
Thermus	COX1/3	552	PNSSFWFFYSAAATLFAFFVAVAA-----LPVNVWVWFLALFAYGLV--R ALED--EYSHPVE---HHTVT	
Oxytricha	COX1	247	-DLNSYIFKIVSYFNVTNVFFLLNYF--LFF--FLEKSFLLFISFSKIFNDLI---FISNFLYKIFYTI--ENF---	
Moneuplo_m	COX1	258	PSLSNSVFRFSFLTGCRIFFFYELVAVILFI--IKSLLFFFLPSTLITSASS---FFSLTWLRITSTVQGL---.. COX1	
Beef mitos	COX3	73	TM3-----TM4-----	
Paracoccus	COX3	80	PAVQKGLRYGMILFIISEVLFFTGFFWAFYHSSLAPT---PELGGCWPTGIHPLNPLEVPLLNTSVLLASGVISITWAHHSMEGTP-----	
Andalucia	COX3	80	PVVRIGLQYGFILFIMSEVMFVAFWFAFIKNALYPMGPDSPKDGWVWPEGLVTFDPWHLPLINTLILLSSGAVTWAHHAFLVLEG-----	
Physarum	COX3	76	--VQTGLRMVLLFIVSEAAFLFAFFWAFHSSSLAPNIEI---GSVWPLGIEPLSAWEVFPFLNTLILLSSGATVWAHHAIVA--GYRKEAILAL	
Thermus	COX1/3	613	--VLLGLRLGFILFIVSEVMFFSFFWAYFHSSLSPNIEI---GSQWPPFALEVIGL--ALPVVNTVILLTSGATITVAHLAILRNKKQIAIESL-	
Oxytricha	COX1	313	-GKSNAWGMWAFIVSEVGLFAILLIAGYLYRLSGAATP-----PEERPALWLALLNTFLLVSSSFTVHFAHH--DLR-----	
Moneuplo_m	COX1	975	PTLTNTIN--FELN--LFSFVFFNMALACTLLIFARA--FFGFSIYFWP--YLCIIFNFLOQLKLFYICKLLKIFSTSFLEFQNYFF-- [7]	
			PVYGKKLI--FENILIFSVLLFNDFLLQLFIVLLV--FFLL-----WDSFLLIITAVCFLSVVALFAWLLDADIYINFLII [14]	
Beef mitos	COX3	155	TM5-----TM6-----	negative side
Paracoccus	COX3	167	---DRKHMLQALFITITLGVYFTLLQAS--EYVEAP---FTISDGVYGSTFVATGFH--GLHVIIGSTFLIVCFPRQL--KFH-----	
Andalucia	COX3	168	---DRKTTINGLIVAVILGVCFGLQAY--EYSHA---FGLADTVYAGAFYMATGFH--GAHVIIGTFLFVCLIRLL--KGQM-----	
Physarum	COX3	165	---ILTISLAAVFTALQIFEYATASFSLSDGIYGS-----TFMATGFH---GFH--VFVGTCAVTCVLLRQIRYH-----	
Thermus	COX1/3	682	--IATIVLALVFTAIQMYEYRHAFPSISDGIYGS-----VF---YMLTGFH---GIH--VLIGTIFICVQFVRLTKDH-----	
Oxytricha	COX1	405	--RRGRFNPFRFGLLVITILGVLFVLVQS--EYVQ-----FYHH--SSWQENLWTAAPFTIVGLHGLHVVIGFGLLILAYLQALRGK----	
Moneuplo_m	COX1	1064	FFSNKNFLYNYFTFLSYFNSFFKYLGNESKFFDPDLK--RFDINDFANNFGLNSYNIFKISTNNLYAVGSKTNIIRYMFKS-----	
			FISLFPREAREGSNRFLIFSF--FGLFAGIFSCFSNVETVVIIGLPFL--VTYYDWGIFNFYFTDLQLLSDIYVVLGAGLFIILMNFYLYLVILV	
Beef mitos	COX3	227	negative side TM7-----	
Paracoccus	COX3	240	-----FTSNHFCFEAAA--WYWHFVDVWVLFVYSIYWWGS-----	C terminus 260
Andalucia	COX3	232	-----TQKOHVGFEEAAA--WYWHFVDVWVLFVVIYIWR-----	C terminus 273
Physarum	COX3	227	-----FTTSHHFCFEAAA--WYWHFVDVWVLFVSIYWWGQ-----	C terminus 267
Thermus	COX1/3	761	-----LLSNHLCFEACA--WYWHFVDVWVLLFVIVYAYGSNAL----	C terminus 264
Oxytricha	COX1	477	-----ITLHNCLEAASMYWHLVDVWVIVITIFVW-----	C terminus 791
Moneuplo_m	COX1	1197	MCSRNNYVEYPSYWIYNTVNGTDVYGRKFTKEYSDNLLSKLSW--SNLIYIEN.. COX1 & COX4? ---	1331
			IYCLRASQKRSDFFITTFDNTVVRGNFMRTQDVQSQILTRATVVRVWQKKS	C terminus 1203

B



C



**Figure 4. Analysis of the molecular architecture of COX3 in bacteria and protists. A – Alignment of bacterial and mitochondrial COX3 proteins.** A set of aligned COX3 sequences from bacteria and protists was initially obtained from the DELTBLAST option of multiple alignment and subsequently implemented manually following data available from the structure of beef [59,60], *Paracoccus* [61] and *Thermus* [54] *aa<sub>3</sub>* oxidase. Residues that bind phospholipids with either H or  $\pi$  bonds [60] are in yellow character and highlighted in dark grey, while those conserved are in bold character. Light grey areas indicate transmembrane helices (TM). **B – Graphical representation of COX1-3 fused proteins.** The hydrophobic peaks in the hydropathy profile of the proteins, which was obtained using the program WHAT [81] with a fixed scanning window of 19 residues, is represented by the sharp triangles, that are commensurated to the peak height (maximum in the hydrophobicity profile) and width of the predicted TM [81], which closely correspond to those observed in 3D-structures [47,54,61]. **C – Deduced sequence of the “minimal” COX operon of protists.** The arrangement of COX genes essentially corresponds to the core sequence of a COX operons of type a (cf. Fig. 3) but in the reverse order of transcription. Dashed symbol represents a protein that may intermix with other COX subunits such as a COX4-like (Fig. S2 in File S1). doi:10.1371/journal.pone.0096566.g004

*NiaD*-related nitrate reductase (Table 1 and Fig. 2B,C). The parallel evolution of mitochondrial sulfite oxidase, which shares the same cytochrome *b<sub>5</sub>* and Moco domains with eukaryotic assimilatory nitrate reductases (Fig. 2B, cf. [38,40]), underlines the intersection of this molecular reconstruction with the evolutionary trajectory of proto-mitochondria.

### 3. Evolution of COX genes and proteins from bacteria to mitochondria

To test alternative evolutionary pathways for mitochondria (Fig. 1B) we next studied the cytochrome *c* oxidase of *aa<sub>3</sub>*-type (also called COX), which appears to be the most common terminal oxidase in extant  $\alpha$  proteobacteria (Fig. 1 and Table S1). In eukaryotes, this enzyme complex is embedded in the inner mitochondrial membrane, combining catalytic subunits of bacterial origin with various nuclear-encoded subunits of unknown function. Although all *aa<sub>3</sub>*-type oxidases are of type A according to the classification of heme-copper oxygen reductases [26], the complexity of their gene clusters has not been considered before. Here, we have analysed in depth this complexity for it provides valuable phylogenetic information. Various aspects of our analysis are presented below in the following order: 1, diversity of COX operons; 2, evolution of COX operons; 3, possible COX operons of proto-mitochondria; 4, evolution of the molecular architecture of COX3; 5, phylogenetic distribution of COX operons.

**3.1 Diversity of COX operons.** We have initially undertaken a systematic analysis of the genomic diversity of *aa<sub>3</sub>*-type oxidases. The scrutiny of all the gene clusters containing proteobacterial COX subunits [46–51] suggests that they fall into three distinctive types of COX operons, which we called type a, b and a–b transition (Fig. 3A – see Table S2 and “Classification of bacterial COX operons” in File S1 for a detailed account of this classification). COX operon type a is divided in four subtypes on the basis of COX1 length and diverse adjacent genes (Fig. 3). These subtypes form coherent clades in the phylogenetic trees of their COX1 subunit (Fig. 3B). Despite the variation in gene sequence, all COX operons appear to derive from the core structure of the *ctaA-G* operon of *Bacillus subtilis* [46–51] (Fig. 3A), which consists of the catalytic subunits *ctaC* and *ctaD* (corresponding to mitochondrial COX2 and COX1, respectively) followed by the hydrophobic, non-catalytic subunit *ctaE* (corresponding to mitochondrial COX3) and *ctaF* (also called COXIV or COX4). Mitochondrial DNA (mtDNA) of eukaryotes generally encodes for COX1, COX2 and COX3 [48]. In bacteria, these principal subunits are often combined with proteins for the assembly of the metal cofactors of the oxidase: *ctaA* (heme A syntase or COX15), *ctaB* (protoporphyrin IX farnesyl transferase, or COX10) and *ctaG* (Cu-delivery protein, or COX11).

Our systematic analysis of bacterial COX subunits has revealed a novel fusion between COX1 and *ctaF*/COX4 (Fig. S2 in File S1). This fusion appears to be restricted to COX operon type a-II (Table S2 in File S1 and Fig. 3A) that often contains Pyrroloquinoline quinone (PQQ)-dependent dehydrogenases such as methanol dehydrogenase related to *MxaF* (Fig. 3A). COX4 is broadly related

to the *ctaF* subunit, which is the least conserved in the *caa<sub>3</sub>*-type oxidase of *Thermus* and *Bacillus* [47] but can be recognized as part of Cyt\_c\_ox\_IV (pfam12270 [52]). However, the diverse forms of short hypothetical proteins that intermix with COX subunits (Fig. 3A) are generally not recognized as members of this family in BLAST searches, due to the wide variation in their size and sequence [47]. Therefore, we have developed a method that quantifies the sequence similarity with the COXIV proteins from *Rhodobacter* [53] and *Thermus* [47,54], for which the 3D structure is available (see Fig. S2 in File S1 and its legend for details). Strong sequence similarity with these COX4 proteins was found in the C-terminal extension of bacterial COX1 proteins that are 630 to 670 aa long, as well as in mitochondrial COX1 of the pathogenic fungus, *Zygomoseptoria tritici* [55] (Fig. S2A in File S1). We additionally identified the sequence signatures of COX4 in small proteins previously recognized as domain with unknown function (DUF [52]) families, namely DUF2909 and DUF983 (Figs. 3 and S3 in File S1). Moreover, the C-terminal part of the mtDNA-encoded COX1 of ciliates, an ancient and diverse phylum of unicellular eukaryotes [56], shows some sequence similarity encompassing both transmembrane helices of COX4 proteins (Fig. S2A and B in File S1). Although this similarity is clearly weaker than that observed with bacterial COX1 proteins, it lies in a conserved region among ciliates (Fig. S2A in File S1 and data not shown) thereby suggesting that fusion of COX1 with COX4 might represent an additional trait shared by bacteria and mitochondria.

**3.2 Evolution of COX operons.** The identification of COX4-like proteins has been combined with phylogenetic analysis to deduce the possible evolution of COX operons. The long proteins derived from the fusion of COX1 with COX3 (hereafter called COX1-3) seem to be the most distant from their mitochondrial homologues (Fig. 3B). These proteins are characteristic of *caa<sub>3</sub>* oxidases [46,47], as well as of COX operon type a, which can therefore be considered the ancestral form of proteobacterial gene clusters for *aa<sub>3</sub>*-type oxidases (Fig. 3A). The differentiation into other types of COX operons can be evaluated also from the phylogenetic trees of the catalytic subunit COX1, the analysis of which has offered new evidence for discriminating the evolutionary pathways in Fig. 1B.

COX1 proteins fused with COX4 (see above) appear to follow the ancestral COX1-3 in phylogenetic trees and are always upstream of a major bifurcation in two large groups: one containing only proteins of COX operon a-b transition that are present in  $\beta$ - and  $\gamma$ -proteobacteria, and the other containing bacterial COX1 proteins of COX operon type b together with their mitochondrial homologues (blue square in Fig. 3B). Mitochondrial COX1 proteins cluster in a monophyletic clade that lies in sister position of closely packed bacterial sub-branches, especially that containing Rhodospirillales (Fig. 3B). This overall tree topology is consistently found with all methods, whereas the branching order within the group containing the mitochondrial clade may vary, depending upon the method and taxa used to construct the phylogenetic trees (Fig. 3B and data not shown). Nevertheless, it is noteworthy that all the



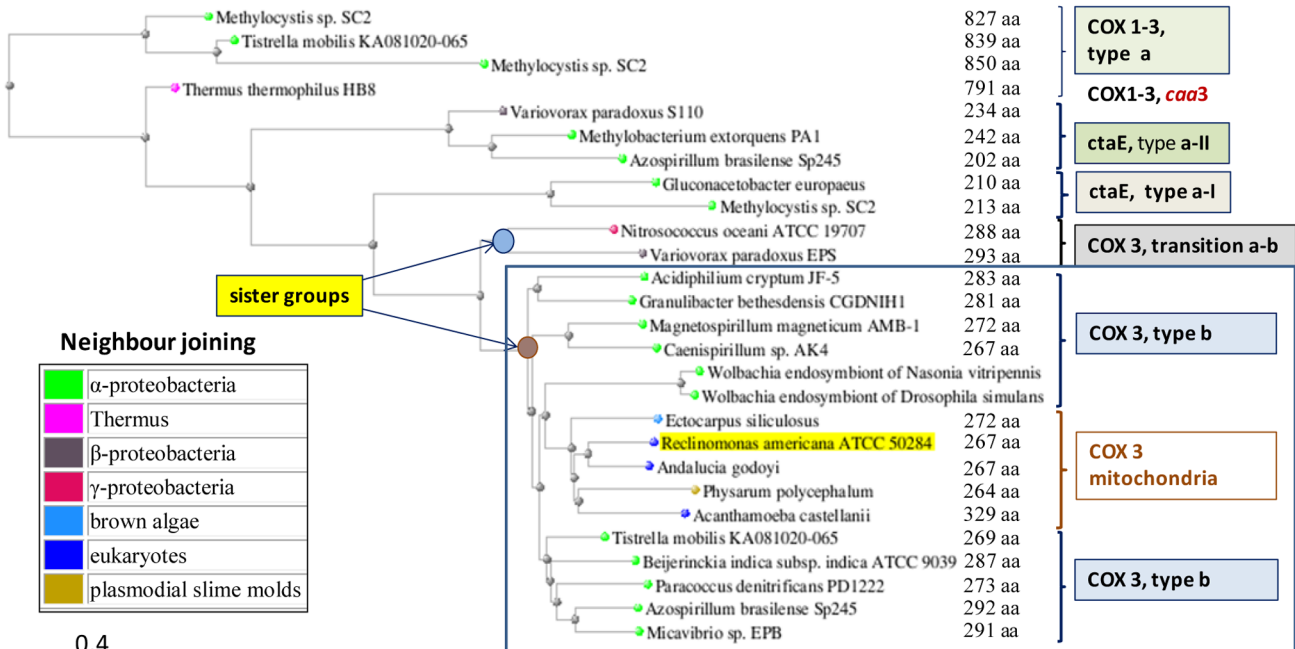
A

Organism and protein	protein	COX operon	length aa	spans TM	PL binding sites				
					E90	PG1	PG2	PE2	PE1
<i>Methylocystis</i> sp. SC2	COX1-3	a	827	6(7)		2 sites	0 site	1 site	
<i>Rhodo_palu_BisA53</i>	COX1-3	a	841	6		1 site	1 site	2 sites	
<i>Methylocystis</i> sp. SC2	COX3a	a-III	236	6		2 sites	3 sites	2 sites	
<i>Variovorax_par</i> EPS	COX3b	a-III	234	6	D	2 sites	>3 sites	2 sites	
<i>Roseomonas</i> sp. B5	COX3	a-I	217	5	yes	2 sites	>3 sites	3 sites	
<i>Gluconacetobacter_europ</i>	COX3	a-I	210	5	yes	3 sites	>3 sites	3 sites	
<i>Variovorax_par</i> S110	COX3	a-II	201	5	yes	2 sites	3 sites	2 sites	
<i>Methylocystis</i> sp. SC2	COX3	a-II	242	5	yes	2 sites	3 sites	2 sites	
<i>Methylophi_bac</i>	COX3	transition a-b	299	7	yes	>3 sites	3 sites	2 sites	
<i>Nitrosococcus_oceani</i>	COX3	transition a-b	288	7	yes	3 sites	>3 sites	>3 sites	
mitochondria, alveolata									
<i>Plasmodium_yoelii</i>	COX3	mtDNA	263	6	yes	2 sites	3 sites	1 site	
<i>Monoeuplotes_minuta</i>	COX1-3	mtDNA	1203	2&5	S	1 site	3 sites	>3 sites	
<i>Oxytricha_trifallax</i>	COX1-3	mtDNA	1331	6(7)	S	2 sites	>3 sites	2 sites	
mitochondria									
<i>Beijerinckia_indica</i>	COX3	b	287	7	yes	>3 sites	>3 sites	2 sites	
<i>Caenispirillum</i>	COX3	b	275	7	yes	>3 sites	>3 sites	>3 sites	
<i>Wolbachia_Dro_sim</i>	COX3	b broken	275	7	yes	>3 sites	>3 sites	>3 sites	
<i>Paracoccus</i>	COX3	b broken	273	7	yes	>3 sites	>3 sites	>3 sites	
mitochondria									
<i>Dictyostelium</i>	COX3	mtDNA	435	7	yes	3 sites	2 sites	3 sites	
<i>Acanthamoeba</i>	COX3	mtDNA	329	7	yes	>3 sites	>3 sites	3 sites	
<i>Andalucia</i>	COX3	mtDNA	267	7	yes	>3 sites	>3 sites	3 sites	
<i>Bos taurus</i> (beef)	COX3	mtDNA	260	7	yes	>3 sites	>3 sites	>3 sites	

Legend

binding	
conserved sites	strength
>3 sites	strong
3 sites	weak
2 sites	very weak
1 or 0 site	none

B



**Figure 5. Structure-function features of COX3 gradually evolved from bacteria to mitochondria. A – Heatmap for the strength of phospholipid binding by COX3 proteins.** The table summarises the molecular features of PL-binding sites (residues) in aligned COX 3 proteins (Table S4 in File S1); it is colour mapped according to the number of conserved sites to represent the increasing PL-binding strength along bacterial and mitochondrial protein sequences, as indicated by the legend on the right of the table. PL-binding is considered weak when less than 3 sites are

conserved for each PL, the nomenclature of which is taken from Ref. [60]. PE, phosphatidyl-ethanolamine; PG, phosphatidyl-glycerol. The list includes conserved amino acids corresponding to E90 in beef *COX3*, which lies near bound PL modulating oxygen entry into the catalytic site of the oxidase [60]. Abbreviations for organisms are: *Rhodo\_palu\_BisA53*, *R. palustris* BisA53; *Variovorax\_par*, *Variovorax paradoxus*; *Methylophi\_bac*, *Methylophilales bacterium* HTCC2181; *Wolbachia\_Dro\_sim*, *Wolbachia* endosymbiont of *Drosophila simulans*. **B - Representative distance tree of COX 3 proteins.** The tree was obtained as described in the legend of Fig. 3B, using as a query the C-terminal region of the *COX1-3* protein from *R. palustris* BisA53 (Accession: YP\_782773, residues 550 to 841) that aligns with bacterial and mitochondrial *COX3* (Fig. 4A). The group containing bacterial proteins from *COX* operon type b and their mitochondrial homologues is enclosed in a blue square as in Fig. 3B. doi:10.1371/journal.pone.0096566.g005

proteins belonging to *COX* operon type b lie in the same group containing the mitochondrial clade, as exemplified in Fig. 3C. Hence, bacteria having only *COX* operon type b cannot be the ancestors of mitochondria. This exclusion encompasses the majority of extant  $\alpha$ -proteobacteria, because the presence of other *COX* operons is restricted to a fraction of these organisms (Table S2 in File S1). We then needed additional information to identify which of the organisms containing multiple *COX* operons may be close to proto-mitochondria. To this end, we next moved to the analysis of *COX* proteins of unicellular eukaryotes.

**3.3. Possible COX operons of proto-mitochondria.** Recently, *COX11* and *COX15* have been found in the mtDNA of Jakobida, an ancient lineage of protists, despite the fact that they are normally coded by nuclear DNA in eukaryotes [48]. The syntenic *COX11COX3*, as well as that of *COX1* adjacent to *COX2* (Fig. 3A), may be considered a relic of bacterial operons that has been retained in the mtDNA of eukaryotes [48]. Are these cues pointing to the original *COX* operon(s) of proto-mitochondria?

To answer this question, we searched the available mtDNA genomes of unicellular eukaryotes. Mitochondrial DNA normally contains separate genes for *COX1*, *COX2* and *COX3* [48] except for aerobic ciliates, in which *COX3* appears to be missing [56,58]. However, we have recognized the sequence signatures of the *COX3* protein within the very long *COX1* of the hypotrichous ciliate, *Oxytricha* [56] (Fig. 4). The *COX1* protein of another hypotrich, *Monoeploptes minuta* [58], appears to contain a split version of *COX3* having its initial two transmembrane helices separated from the subsequent 5-transmembrane helices domain by the major part of *COX1* (Fig. 4). The mtDNA of ciliates often contains split genes [56,58], but in this case an ancestral splitting of *COX3* must have been subsequently intermixed with the *COX1* gene. The alternative possibility would be that *COX3* splitting may reflect a fusion between precursors of mitochondrial *COX3*, since in *Monoeploptes* it occurs within the region joining the two transmembrane domains which form the V-shaped structure of the protein [53,59-61].

In any case, the novel identification of a *COX3*-like protein embedded within the long *COX1* gene of unicellular eukaryotes (Fig. 4) suggests that the primordial form of such a chimaeric gene was a *COX1-3* protein equivalent to those of bacterial *COX* operons of type a. By considering the gene order in ciliate mtDNA [56,58], we have deduced the possible sequence of the “minimal” *COX* operon that might have been present in the ancestors of ciliate mitochondria (Fig. 4C). The gene sequence closely resembles the core structure of a *COX* operon of type a - in the opposite order of transcription (cf. Fig. 3A and 4C) - and is clearly different from the sequence of *COX* operon type b (Figs. 3A and S3 in File S1). In view of the consensus that a single event of symbiosis originated all mitochondria [1-10] and considering the presence of *COX11COX3* synteny in Jakobide mitochondria [48], a feature characteristic of *COX* operon type b (Figs. 3 and S3 in File S1), we surmise that proto-mitochondria possessed two different *COX* operons: one of type a and another of type b. Differential loss of either operon might further explain some differences in the mtDNA-coded proteins of ciliates and other unicellular

eukaryotes, as well as the different types of accessory subunits of their bioenergetic complexes [1]. Of note, phenetic analysis sustains the similarity between the *COX* gene sequence of protists and bacterial *COX* operon of type a-II, in particular those lacking an isolated *COX4* as in *Methylobacterium extorquens* PA1 (Table S3 in File S1).

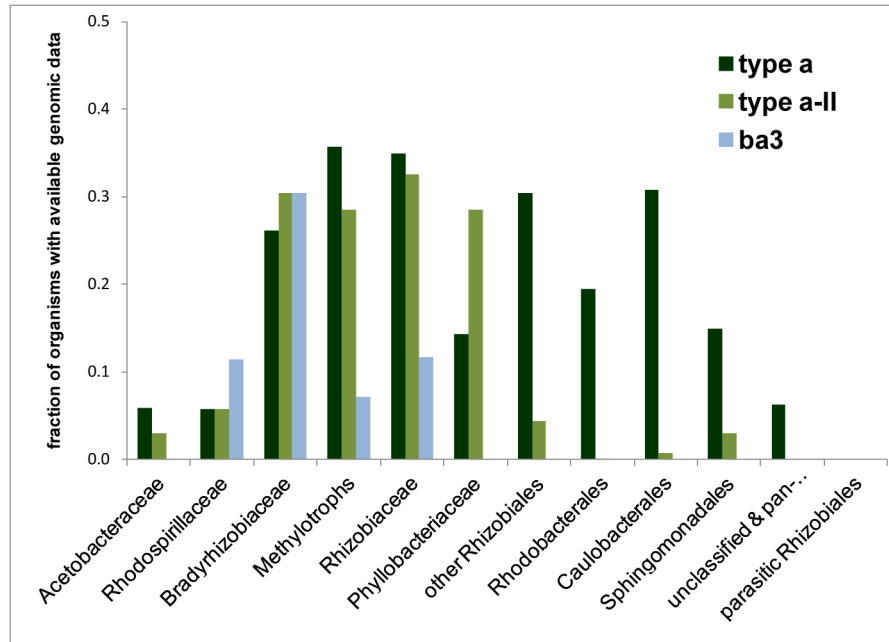
**3.4 Evolution of the molecular architecture of COX3.** In the 3D structures available for cytochrome *c* oxidases, the initial two transmembrane helices of the 7-helices *COX3* protein that is present in mitochondria and bacterial *COX* operon type b (Fig. 3A) are involved in the binding to membrane phospholipids (PL) [53,59-61]. The tight binding of two specific forms of these PL to mitochondrial *COX3* appears to modulate the entry of oxygen into the binuclear catalytic centre of the enzyme [60]. PL-binding residues are present also in other parts of the *COX3* protein that are common to all its forms and tend to be conserved [59-62]. Here, we have evaluated the amino acid substitutions of the PL-binding sites in *COX3* (Table S4 in File S1) by translating residue variation into PL-binding strength (Fig. 5A). The results of this analysis are consistent with the phylogenetic trees of *COX3*, in which a major bifurcation separates the  $\beta$ - and  $\gamma$ -proteobacterial proteins from those of  $\alpha$ -proteobacteria that are grouped together with mitochondrial *COX3* (Fig. 5B). The overall tree topology of *COX3* proteins thus matches that of *COX1* proteins, even if the internal branching of  $\alpha$ -bacteria with the mitochondrial clade appears to be different (Fig. 5B cf. Fig. 3B).

Quantitative evaluation of the PL-binding strength further refines the evolutionary relationship among *COX3* proteins. First, it shows that the 5-helices form of the protein belonging to *COX* operon type a-II occupies an intermediate position between ancestral *COX1-3* and the 7-transmembrane form of *COX3* (Fig. 5A). Secondly, it allows the comparison with the highly divergent sequence of ciliate *COX3* embedded within *COX1* (Fig. 4), which shows a PL-binding strength lying mid-way between that of *COX3* proteins of type a-II operon and those of other protists (Fig. 5A and Table S4 in File S1). Finally, bacterial *COX3* of *COX* operon type b has essentially the same PL-binding strength as that of mitochondrial *COX3* (Fig. 5A and Table S4 in File S1), thereby weakening the structural and phylogenetic significance of variable inter-group branching between  $\alpha$ -bacterial and mitochondrial *COX3* sequences (Fig. 5B and data not shown).

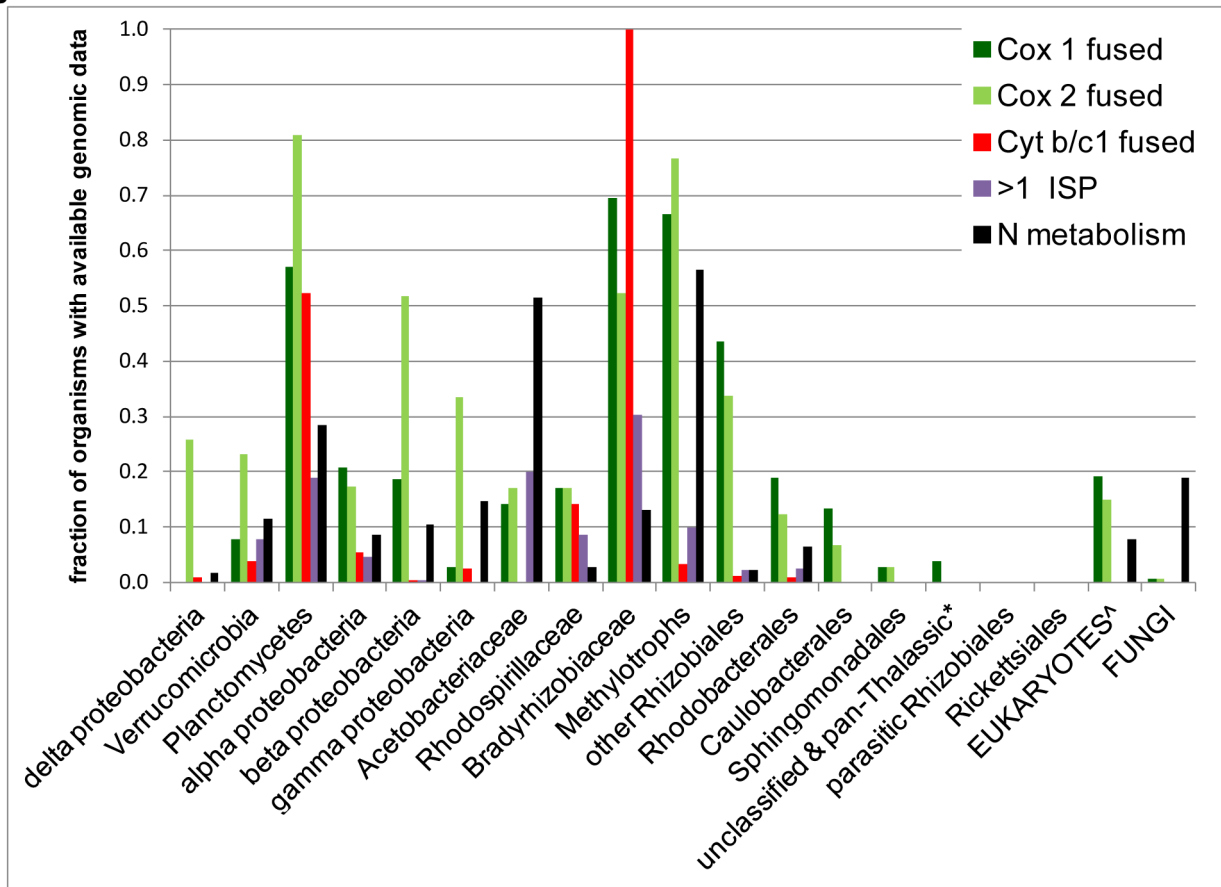
**3.5. Phylogenetic distribution of COX operons.** To acquire further information for differentiating the pathways of mitochondrial evolution in Fig. 1B, we studied the phylogenetic distribution of diverse *COX* operons. The vast majority of Rhodobacterales, Sphingomonadales and Caulobacterales, together with unclassified  $\alpha$ -proteobacteria such as *Micavibrio* and the SAR11 clade - which we include here under the generic label of ‘pan-Thalassic’ - possess only *COX* operons of type b. This implies that *Roseobacter* and *Micavibrio* cannot be related to the ancestors of mitochondria, as for *Pelagibacter* and similar marine organisms.

On the other hand, 40  $\alpha$ -proteobacterial organisms and several  $\beta$ -proteobacteria combine *COX* operon type b with a type a-II operon, the phylogenetic distribution of which is similar to that of *ba3* oxidases [26] (Fig. 6A). Conversely, *COX* operon type a-I has the broadest phylogenetic distribution among all types of *COX*

A



B



**Figure 6. Taxonomic distribution of bioenergetic systems in bacteria. A - Distribution of COX operon types in major families of  $\alpha$ -proteobacteria.** The frequency of each type of COX operon was normalised to the number of  $\alpha$ -proteobacterial organisms with genomic data that are currently available (from NCBI resources <http://www.ncbi.nlm.nih.gov/taxonomy/> - accessed 14 March 2014) [50]. See Table S2 in File S1 for a detailed list of the taxonomic distribution of diverse COX operon types. The definition 'pan-Thalassic' collects together organisms of the SAR clade

with *Magnetococcus*, *Pelagibacter* and *Micavibrio*. **B. -Distribution of fused proteins and N-metabolism elements along diverse bacterial lineages.** Fused proteins were identified with the combined resources of NCBI and the Protein Family website (PFAM 27.0 - <http://pfam.sanger.ac.uk/> [52]). Multiple forms of ISP were counted as >1 ISP. Taxa are arranged according to their approximate phylogenetic position considering also metabolic features (cf. Refs [5,31]). For each group, the frequency is normalized as in A. Eukaryotes (°) include amoebzoa, ciliates and heterokonts. N-metabolism encompasses: methane monooxygenase, ammonia monooxygenase, nitrite oxidoreductase, *Nirf* nitrite reductase and its homologues in COX operon type a-I (Fig. 3A), ammonia oxidation and anaerobic ammonia fermentation [30,32]. doi:10.1371/journal.pone.0096566.g006

operon, encompassing taxonomic groups beyond the *phylum* of proteobacteria such as Planctomycetes [50]. Indeed, the *Nif*-like gene cluster that is associated with this *COX* operon was originally discovered in ancient eubacteria including Planctomycetes [63]. Although the functional implications of the combination of a *Nif*-like operon with a *COX* gene cluster remains unknown, we are intrigued by the possibility that the overall gene sequence would produce a compact electron transport chain from quinol, or products of N metabolism, to oxygen [32,50]. Consequently, *COX* operon type a-I would represent the ultimate bioenergetic connection between cytochrome *c* oxidase and N metabolism, a fundamental concept in our approach to discern mitochondrial evolution (Fig. 1).

#### 4. Phylogenetic distribution of N metabolism and fused proteins in bacteria and mitochondria

To explore the phylogenetic dimension of the connection between *COX* operons and elements of N metabolism, we studied the taxonomic distribution of *NifD* and other key elements of the N cycle in conjunction with that of fused subunits of *aa<sub>3</sub>*-type oxidases (Fig. 6B). Indeed, *COX* operon type a-I invariably contains *COX2* fused with a *c*-type cytochrome (Figs. 3A), a fusion which is frequently present also in other *COX* operons (Fig. 3A and Fig. S3 in File S1). Fusion between catalytic subunits of bacterial heme-copper oxidases has been noted before [47,64], but considered a nuisance for phylogenetic analyses [64]. However, it constitutes a relic of ancestral bacteria adapted to harsh conditions in which the compact structure of bioenergetic systems would have been advantageous [47]. Since we have now shown that fusion between *COX* subunits is present also in the mitochondria of unicellular eukaryotes (Fig. 4) and fungi such as *Phaeosphaera* [57], we could consider them as potential relics of the evolutionary past of mitochondrial bioenergetics.

We therefore evaluated the frequency and phylogenetic distribution of fused *COX* subunits and also of the fused proteins that are present in the cytochrome *bc<sub>1</sub>* complex, the cytochrome *b* subunit of which has been previously reported to be fused with the cytochrome *c<sub>1</sub>* subunit in *Bradyrhizobium* [65]. We found the same fusion in all members of the Bradyrhizobiaceae family plus some Rhodospirillales (Fig. 6B), as well as in Planctomycetes [66].  $\alpha$ -proteobacteria show the highest frequency of fused cytochrome *b* among proteobacterial lineages, thereby suggesting that this type of protein was present before the separation of  $\beta$ - and  $\gamma$ -proteobacteria. Conversely, many more  $\beta$ -proteobacteria possess fused *COX2* proteins than  $\alpha$ -proteobacteria (Fig. 6B).

Within  $\alpha$ -proteobacteria, the distribution of fused *COX* and cytochrome *b* proteins follows a bell-shape profile along the likely evolutionary sequence of the taxonomic groups (Fig. 6B, cf. [5]). Some Sphingomonadales and Caulobacterales have fused *COX* proteins without possessing bioenergetic elements of N-metabolism (Fig. 6B). Parasitic Rhizobiales, Rickettsiales and pan-Thalassic organisms lack both fused bioenergetic proteins and elements of N-metabolism, in contrast with amoebzoa, fungi and heterokonts (Fig. 6B cf. Table 1). The absence of the above characters in parasitic and pan-Thalassic organisms could derive from their highly streamlined genomes. However, the high frequency of fused

genes in other taxa does not correlate with genome size, since acetic acid bacteria, which have a comparatively small genome, show a higher frequency of fused *COX2* proteins than, for instance, Rhodobacteriales (Fig. 6). Our interpretation of the data presented in Fig. 6 is that fused bioenergetic proteins and elements of N metabolism are preserved together in phylogenetically ancient groups of  $\alpha$ -proteobacteria, from which they have been passed to proto-mitochondria but then progressively lost along the differentiation of other  $\alpha$ -proteobacteria. This implies that Methylotrophs, Bradyrhizobiaceae and several Rhodospirillales would be the oldest extant organisms of the  $\alpha$ -proteobacterial lineage, and consequently close to the distal progenitors of proto-mitochondria.

The phylogenetic distribution and similar genomic arrangement of fused bioenergetic proteins (Fig. 6) raises the question as to whether they may derive from events of Lateral Gene Transfer (LGT), for example with Planctomycetes [67]. However, detailed analysis of the molecular architecture of cytochrome *b* proteins (M. Degli Esposti, unpublished data) and the overall consistency of distance trees of fused proteins with established phylogenetic relationships (Fig. 3) indicate that LGT events have minimally contributed to the observed distribution of fused bioenergetic proteins and their diverse genomic clusters.

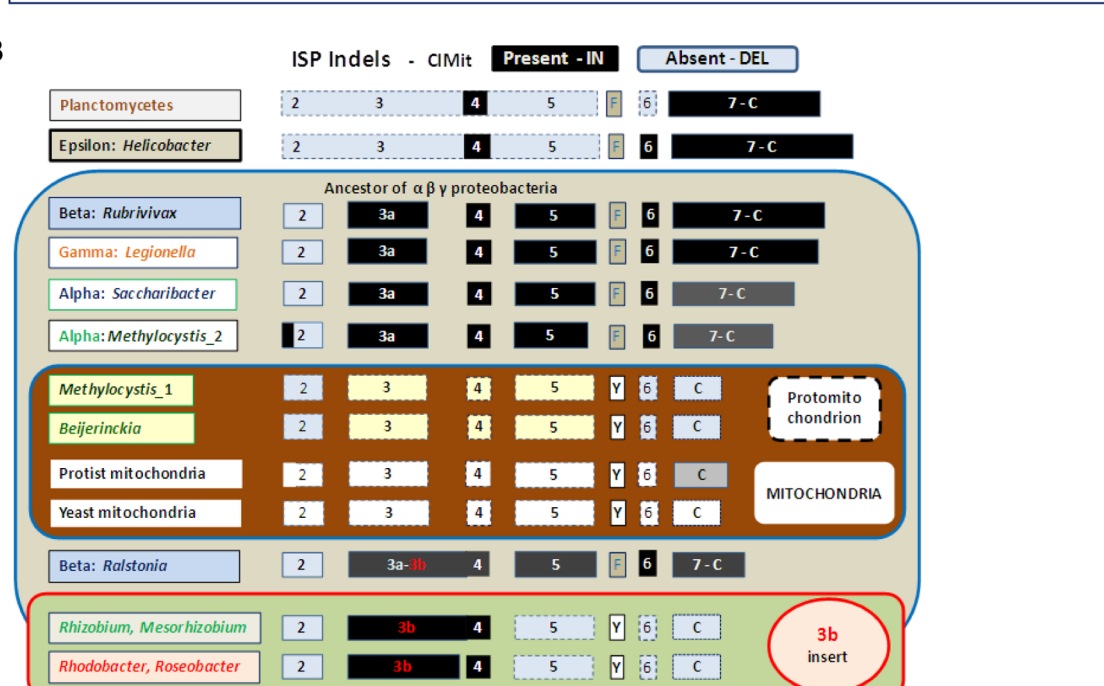
#### 5. A complementary approach: the molecular evolution of nuclear encoded ISP

To complement the above analysis of mtDNA-encoded proteins of the *aa<sub>3</sub>*-type oxidase, we next examined the molecular evolution of the “Rieske” iron sulfur subunit (ISP) of the cytochrome *bc<sub>1</sub>* complex. This ubiquitous redox protein is coded by the nuclear DNA and therefore does not suffer from the distortions due to the fast mutation rate of mtDNA-encoded proteins [16,37,48]. Its precursor form, once imported into mitochondria, matures within the intermembrane space where its catalytic core resides. After implementing structure-based alignments (Fig. S4 in File S1), we noted diverse insertions that are present in the catalytic core of ISP proteins from different lineages, which we have named CIMit - Conserved Indels vs. Mitochondria (Fig. 7 and Fig. S4 in File S1). CIMit3 is the most prominent of these insertions, lying at the surface of bacterial *bc<sub>1</sub>* complexes [68,69] with parallel inserts in the partner protein, cytochrome *b* [68–72]. This and other indels (according to the definition in Ref. [73]) seem to carry valuable phylogenetic information, enabling the resolution of relationships that are blurred in phylogenetic trees (cf. Figs. 7C and 8). For instance, only *Tistrella* ISP has no residues corresponding to the CIMit5 insertion among the proteins from Rhodospirillaceae (Fig. 7), while in distance trees these proteins appear to be equally close within a sister sub-branch of their mitochondrial homologues (Fig. 8).

*Methylocystis* sp. SC2 and a few other Rhizobiales have a second, longer ISP (ISP2) that resembles the proteins from acetic acid,  $\beta$ - and  $\gamma$ -proteobacteria, with which it clusters together in distance trees (Figs. 7, 8 and S4 in File S1). Contrary to the latter organisms, ISP2 is not present within the *petABC* operon of the *bc<sub>1</sub>* complex but in isolated gene clusters that have no common flanking genes (not shown). Hence, ISP2 may have arisen from gene duplication as reported for the  $\beta$  proteobacterium, *Rubrivivax*

A

WP_020226654	64	PGSLIKTVEW	RGKPVWIMRRTPEMIAALSGQDD	---	KLADP	NSSKDALPE	---	ALRNPGRSERQDLFVAIGICT	130	Acidovorax sp. MR-S7/2
YP_001900760	59	PGQMKVVEW	RGKPVWLLKRTPDMLLESLKKTND	---	EVADP	KSDVPTFMKTPDY	---	CKNENRTRSAEHKDLLVVVIGICS	130	Ralstonia pickettii 123/2
NP_275043	55	AGGLIYAEW	QKGIWILNRTDQQLKDLKGLNG	---	ELTDP	NSDABEQPE	---	YARNENRTRSKPNILVAIGICT	125	Neisseria meningitidis
WP_020564984	59	SGGLIYVW	RGKPVWILNRTPEVLATLQTLDS	---	ELRDP	LSSESIQPS	---	YTKNENRTRSKPEIFVAIGICT	125	Methylosarcina fibrata
YP_006985287	73	AGQIIVTW	RGWVFWIQRTEPEMLKTLQDFAIL	---	KLKRD	SESLIFQQPK	---	DATNWHRSVSDIGVMIGICT	141	Gluconobacter oxydans/1
YP_00227526	71	PGQIIVTW	HGSPVFTHRTPESLARLQDEALA	---	ARLDR	PSGALQPE	---	YARNWHRSIKPEGVVIGICT	139	Gluconacetobacter diazo/2
WP_010510616	62	PCQIIVTW	QKGVPTFTRTPESLAALQDAALIT	---	GLRDR	QSNDRQPE	---	YARNWHRSIDPRYGVVIGICT	130	Gluconacetobacter europa/2
CCD32085	59	PGKPVWVW	RGLPVAIFRRPDALEKLRQPVLL	---	ELADP	DSEVLQQR	---	YARNWHRSIDPTVAVLVIGICT	127	Methylocystis sp. SC2/2
YP_577370	69	PCQIIVTW	RHPILVLRNSDAALKTLQDFKLL	---	QLSDP	QSSVFPQPE	---	YARNWHRSANPEYGVVIGICT	137	Nitrobacter hamburgensis/2
WP_019172466	68	PCQIIVTW	RGPVFIINRAADALKEKLNPSLL	---	ARLDR	DSQVQOPE	---	YARNWHRSNPNQVAVLVIGICT	136	Pseudaminobacter salicyl/2
YP_002282274	64	PCNSLTVW	RKGFIFIRNRTPEEVKAADVPILA	---	DLKDP	VARNANLPPEAQAT	---	GVDRSGGKDKENNVIMVIGICT	135	Rhizobium leguminosarum
NP_385925	64	PCQIIVTW	RKGFIFIRNRTDKEVEEAKVALE	---	ELKDP	VARNANLPDAEAS	---	DLDRSAGEGKKNEMIMVIGICT	136	Sinorhizobium meliloti/1
WP_004610841	59	PCNSLTVW	RKGFIVVNRNTEQEMKDEGRAVKLS	---	DLKDP	VARNANLPADAP	---	ATDANRTFGKFAWMMVQVICT	136	Mesorhizobium opportunistum
Rbaphaer 3D	61	EGVQIVTK	KLKPIFIRRT	---	QLQVD	VARNANLDAGAE	---	ALDNLIDEAGEMLVMGICT	132	Rhodobacter sphaeroides
YP_004691057	59	PCQIIVTW	QKGVFIRARTEEEIQAARATDILT	---	DLQDP	LAQANLAGDAD	---	AADENRALSDEGQWLVQMGICT	126	Rhoseobacter litoralis/2
YP_001753157	67	DCQIIVTW	RKGLIFVRKLTAKAVADMKAAPLS	---	AMIDP	---	---	AAFTRTRVKSGHQDWLVVYVIGICT	130	Methylobacterium radio
YP_001640007	59	DCQIIVTW	RKGLIFVRKLTAKAVADMKAAPLS	---	EMIDP	---	---	AAFTRTRVKSGHQDWLVVYVIGICT	118	Methylobacterium_exto_Pa1
YP_002502576	68	ECQIIVTW	RKGLIFVRRLSKEIADMKAAPLS	---	DLDDP	---	---	AFQARVKEGHQDWLVTYGICT	126	Methylobacterium_nodulans
WP_009763700	59	PIAEGQIV	RVFWRKGLIFIRHRTPEEIKAAEDVNV	---	ASLRDP	---	---	QADSARVKGKAQWLVVYVIGICT	121	Microvirga
YP_782883	61	EGQIVTKV	RGKGFIFIRHRTPEEIKAAEDVNV	---	SFRDP	---	---	QADSARVKGKAQWLVVSAICT	120	Rhodopseudo palu BisA53/3
YP_578436	60	EGQIVTKV	RGKGFIFIRHRTPEEIKAAEDVNV	---	SLPDP	---	---	QADSARVKGKAQWLVVYVIGICT	119	Nitrobacter hamburgensis/1
WP_006590199	62	EGQIVTKV	RGKGFIVRHRRTAAEIKSADVDPLS	---	ALPDP	---	---	EPDAKRVQKPEWLVVYVIGICT	120	Methylocystis sp. SC2/1
YP_002364045	60	VGQIIVTW	RKGFIFVSHRTEEEIAAARAVPLN	---	ELKDP	---	---	ATDQSRVQKPEWLVVYVIGICT	117	Methylocella
YP_001831301	59	EGQIVTKV	RGKGFIVRHRRTAAEIKSADVDPLS	---	ELRDP	---	---	QADSARVKGKAQWLVVYVIGICT	116	Beijerinckia indica
WP_019459664	61	AGQIIVTW	RGKGFIFIRHRTPEEIKAAEDVNV	---	ALKDP	---	---	ATDQSRVQKPEWLVVYVIGICT	151	Roseomonas sp. B-5
YP_002299592	69	EGQIVTKV	RGKGFIVRHRRTAAEIKSADVDPLS	---	SLIDP	---	---	QADSARVKGKAQWLVVYVIGICT	128	Rhodospirillum centenum/2
WP_007438995	73	EGQAVTVM	RGKGFIVRHRRTAAEIKSADVDPLS	---	ELRDP	---	---	QADSARVKGKAQWLVVYVIGICT	130	Acetobacter bacter AT-5844
WP_004985728	71	EGQIVTKV	RGKGFIVRHRRTAAEIKSADVDPLS	---	ELRDP	---	---	QADSARVKGKAQWLVVYVIGICT	128	Azospirillum brasilense
YP_423451	66	PCQIIVTW	RKGFIVRHRRTAAEIKSADVDPLS	---	DLRDP	---	---	QADADRACKPEWLVVYVIGICT	123	Magnetospirillum magneticum
WP_009542637	65	EGQIVTKV	RGKGFIVRHRRTAAEIKSADVDPLS	---	ELRDP	---	---	QADADRACKPEWLVVYVIGICT	122	Caenispirillum salinarum/2
WP_008889443	65	VGQIIVTW	RKGFIFIRHRTPEEIKAAEDVNV	---	ELVDP	---	---	QADADRACKPEWLVVYVIGICT	122	Thalassospira profundimaris
YP_006373746	68	AGQAVTVM	RGKGFIVRHRRTAAEIKSADVDPLS	---	DLPDP	---	---	QADADRACKPEWLVVYVIGICT	125	Tistrella mobilis
XP_004356972	116	EGQIIVTW	RKGFIFIRHRTAAEIKSADVDPLS	---	DMRDP	---	---	QADADRACKPEWLVVYVIGICT	175	Acanthamoeba castellanii





**Figure 7. Molecular evolution of the Rieske subunit (ISP) of the cytochrome *bc<sub>1</sub>* complex. A – Alignment of the ISP proteins from bacteria having various *COX* operons.** ISP sequences were selected from the organisms displaying multiple *COX* operons and also ISP forms (Table S2 in File S1 and Fig. 6). The alignment was manually refined using structural information, as detailed in Fig. S4 in File S1. This alignment shows only the catalytic core of the ISP from  $\alpha$ -,  $\beta$ - and  $\gamma$ -proteobacteria, plus *Acanthamoeba* as the sole mitochondrial representative. See Fig. S4 in File S1 for a complementary alignment including the N-terminal transmembrane region and further information, including secondary structure elements (beta sheet in purple and alpha helix in green) and Conserved Indels vs. Mitochondria (CIMit). The accession codes of the proteins are shown on the left of each sequence block, while the organisms are listed on the right abbreviated as follows: *Gluconacetobacter diazo* & *\_europa*, *Gluconacetobacter diazotrophicus* PA1 5 & *europaeus*, respectively; *Pseudaminobacter salicyl*, *Pseudaminobacter salicylatoxidans*; *Methylobacterium radio* & *\_exto\_PA1*, *Methylobacterium radiotolerans* JCM 283 & *extorquens* PA1, respectively; *Rhodospseudo palu\_BisA53*, *R. palustris* BisA53; and *Acetobacter bacter* AT-5844, *Acetobacteraceae bacterium* AT-5844. ISP1 indicates the ISP form that is present in the *petABC* operon. **B – Evolutionary pattern of the conserved indels in bacterial and mitochondrial ISP.** The molecular features deduced by the structure-based alignment of ISP proteins are rendered graphically following the numerical order of conserved indels presented in A and Fig. S4 in File S1. DEletions conserved in bacterial vs. mitochondrial ISP sequences are represented in pale blue boxes with black labels, whereas INserts with respect to mitochondrial sequences are represented in black boxes with white labels.  
doi:10.1371/journal.pone.0096566.g007

*gelatinosus*, where the two forms of the proteins are interchangeable in the complex [74]. The duplicates of *Rubrivivax* ISP are closely related to each other, as in the case of the multiple ISP forms of *Roseobacter* and other Rhodobacterales (Table S1 in File S1). However, ISP2 and the in-operon ISP1 present in the same Rhizobiales organisms are separated by a deep bifurcation in phylogenetic trees, which resembles that seen in *COX1* trees (Fig. 3B,C cf. Fig. 8). Hence, ISP2 is an ancestral character of  $\alpha$ -proteobacteria equivalent to *COX* operons of type a, consistent with their similar phylogenetic distribution (Fig. 6B). Its origin can be traced to the separation of the  $\alpha\beta\gamma$  lineages, probably after the earliest proteobacterial ISP had evolved in a distinct path from its paralogues of the *bcf* complex present in Planctomycetes and Nitrospirales [75] (Fig. 8B). This ancestral form of ISP was in all likelihood devoid of the abovementioned insertions as in ISP1 of *Rhodospseudomonas palustris* BisA53 or *Nitrobacter hamburgensis*, which lie in the most distant branches of phylogenetic trees (Fig. 8A). Of note, these proteins show the single-residue deletion corresponding to CIMit6, which is shared with the ISP proteins of many  $\alpha$ -proteobacteria and their mitochondrial homologues (Figs. 7 and 8).

Importantly, the molecular features of ISP proteins provide crucial information for discriminating between the alternative pathways of mitochondrial bioenergy evolution in Fig. 1B. In particular, bacterial organisms possessing an ISP containing the CIMit3B insert (Figs. 7 and S4 in File S1) can now be excluded from mitochondrial ancestry. This applies not only to Rhodobacterales such as *Roseobacter*, but also to *Rhizobium*, *Sinorhizobium* and *Mesorhizobium* organisms that have *COX* operon type a-II (Table S2 in File S1).

## 6. Analysis of bacteria without *aa<sub>3</sub>*-type cytochrome *c* oxidase

The analysis conducted so far has exploited bioenergetic systems that are not always present together in extant bacteria (Table S1 in File S1). For example, *Magnetococcus* has no functional *aa<sub>3</sub>*-type cytochrome *c* oxidase but a complete operon for the *bc<sub>1</sub>* complex and the *cbb<sub>3</sub>*-type oxidase (Table S1 in File S1, cf. Ref. [76]). Phylogenetic analysis has shown that the sequence of *Magnetococcus* ISP is rather similar to that of protists' mitochondria, even if it shows some unique amino acid changes (Figs. 8B and S4 in File S1). *Magnetococcus* lies in a deep branch of the evolutionary tree of  $\alpha$ -proteobacteria [76], similarly to *Midichloria*, which also has a *cbb<sub>3</sub>*-type oxidase instead of the *aa<sub>3</sub>*-type oxidase of other Rickettsiales [21]. *Midichloria* has an ISP protein with a unique insertion in the conserved cluster-binding region and also an unusually split version of the catalytic, *COX1*-like subunit of *cbb<sub>3</sub>*-type oxidase [21]. These molecular properties seem to indicate a side-path in the phylogenetic relationships with the mitochondrial

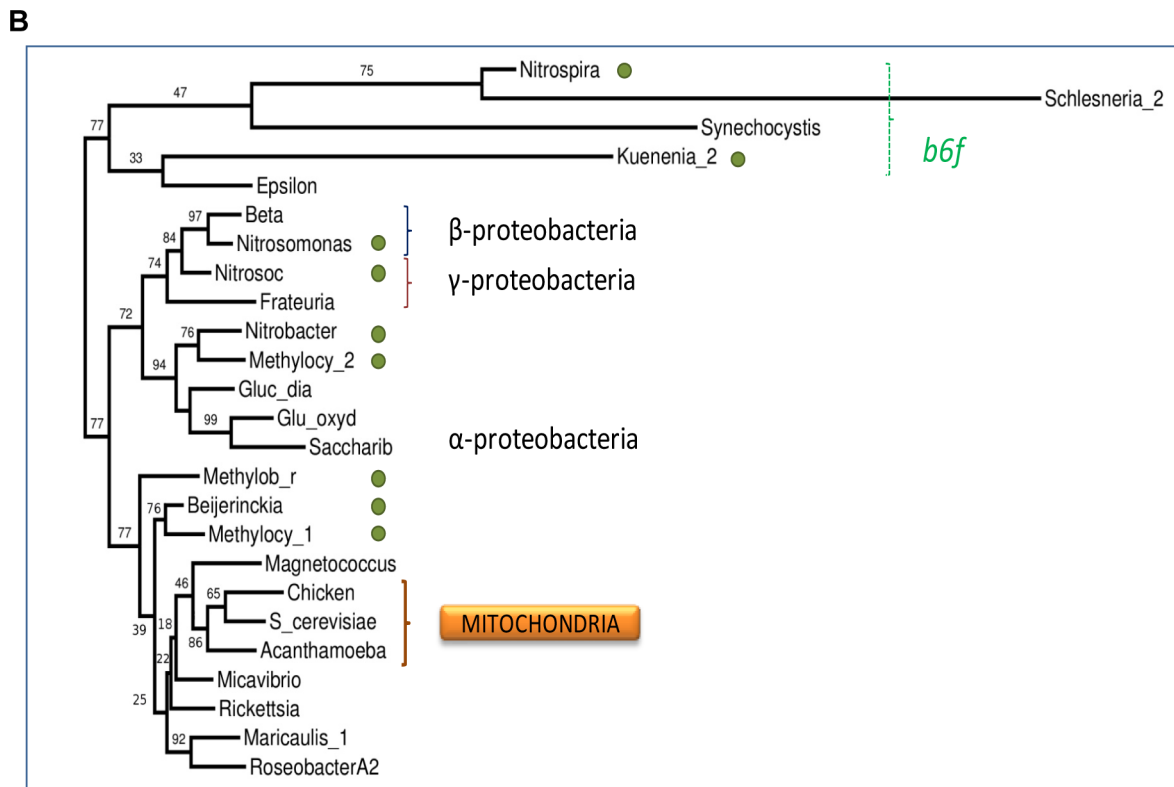
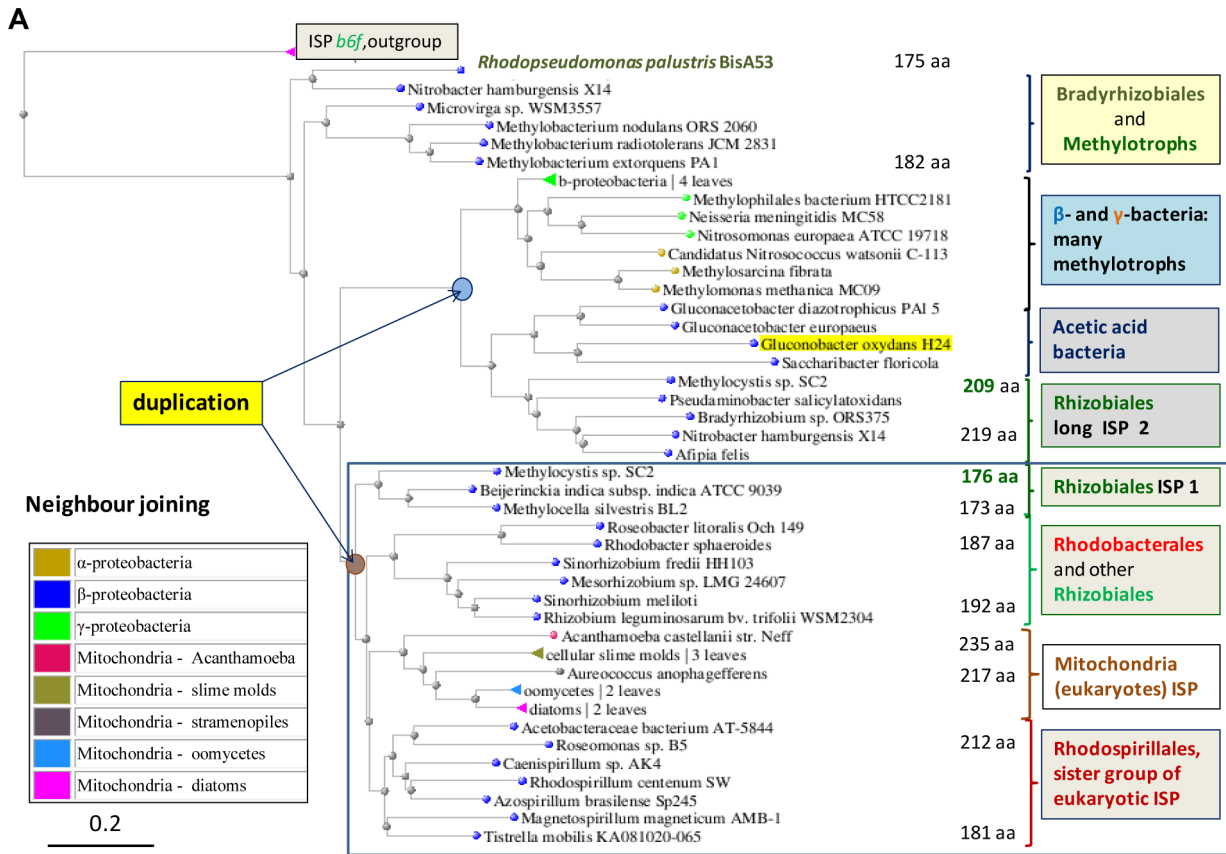
lineage (cf. Fig. 1B), a possibility strengthened by the analysis of the genomic and protein sequences of *cbb<sub>3</sub>*-type oxidase (data not shown). Hence, the scheme in Fig. 1B is consistent with the overall phylogenetic pattern of both *aa<sub>3</sub>*-type and *cbb<sub>3</sub>*-type terminal oxidases.

## Conclusions

Herein, we have followed novel approaches to reconstruct the possible bioenergetic characters of the bacterial ancestors of mitochondria. Rather than taking into consideration all the information that is now available from bacterial and mitochondrial genomes, we have focused on a few proteins that are crucial for bioenergy production in both bacteria and mitochondria and have multiple variants. The diverse molecular forms and genetic organization of bioenergetic systems have been hardly considered in previous studies of phylogenomics; for instance, none of the papers reviewed in Ref. [9] used proteins of energy metabolism. Conversely, recent studies on bacterial oxidases [27,64] have not considered the complexity of *COX* operons (Figs. 3 and S3 in File S1). Here we have classified this complexity and exploited its most informative aspects to reconstruct the molecular evolution of individual protein components that are encoded by either mtDNA or nuclear DNA of eukaryotes. By integrating the information thus obtained, we have excluded that several bacterial lineages previously proposed to be related to mitochondria could be in the direct line of mitochondrial ancestry, in particular the endocellular obligate parasites of the Rickettsiales group and the photosynthetic organisms *Rhodobacter* and *Rhodospirillum*. Our work indicates that mitochondrial ancestors retained bioenergetic elements of N metabolism and the *bd*-type ubiquinol oxidase, which have been subsequently lost in different paths of convergent evolution (Fig. 1B).

In concluding this work, we discuss steps of differential loss also in conjunction with the possible acquisition of systems or proteins via LGT, to provide a complete account of the remaining possibilities for the evolution of mitochondrial bioenergy production (Figure 9). Multiple lines of evidence emerging from our work lead to the conclusion that the subset of bioenergetic systems lacking the *cbb<sub>3</sub>*-type oxidase - typical of methylotrophs and *Gluconacetobacter* (Table S1 in File S1) - probably matches the bioenergetic capacity of the distal ancestors of mitochondria. This evidence includes the maximal diversity of *COX* operons and N metabolism in the abovementioned organisms (Tables S1 and S2 in File S1). The ancestral organisms from which proto-mitochondria emerged in all likelihood evolved just after the separation of  $\beta$ - and  $\gamma$ -proteobacterial lineages, a concept that is sustained, in particular, by the taxonomic distribution of fused bioenergetic proteins and key elements of N metabolism (Fig. 6). At the whole taxon level,  $\beta$ - and  $\gamma$ -proteobacteria have a much higher





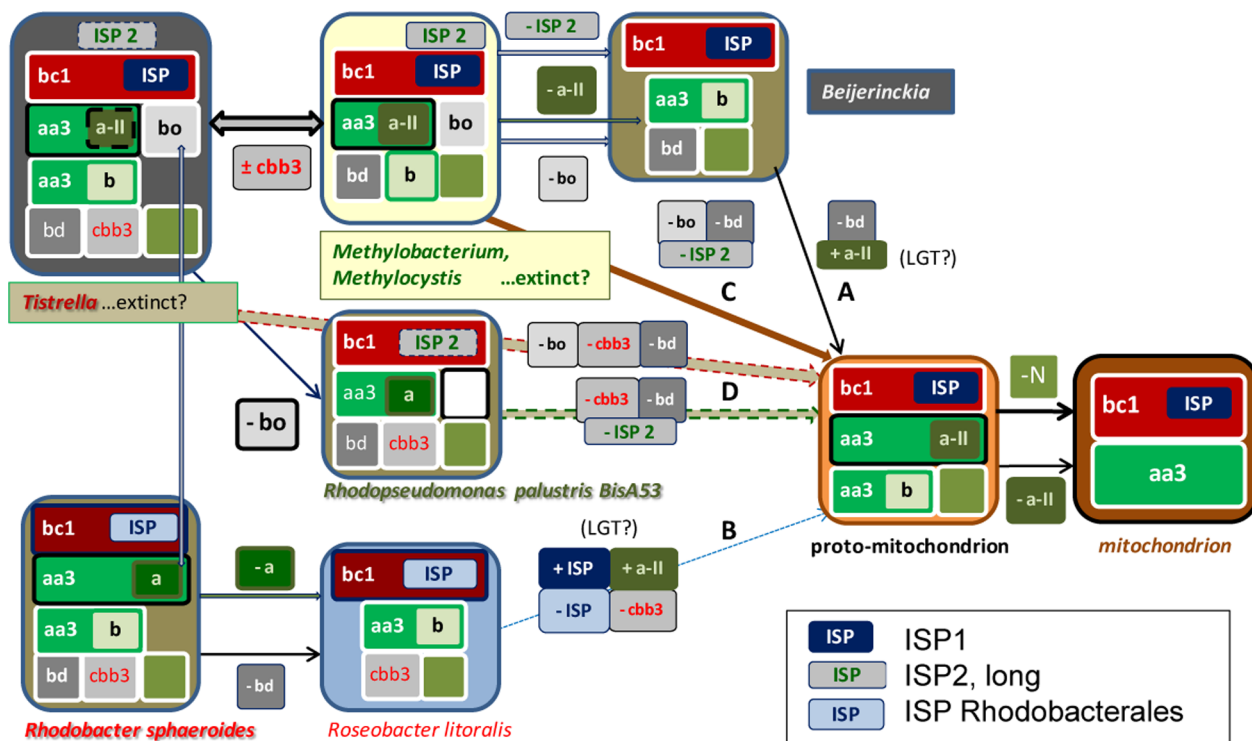
**Figure 8. Phylogenetic relationships between diverse forms of ISP. A – Distance tree encompassing proteobacteria and mitochondria.** The tree was obtained as described in the legend of Fig. 3B using the alignment of Fig. 7A and two ISP proteins from the *b<sub>6f</sub>* complex as outgroup (top). The group containing bacterial ISP1 proteins together with their mitochondrial homologs is enclosed in the blue square to

highlight a likely ancestral duplication separating it from the group with ISP2. **B – Long distance phylogenetic relationships of bacterial ISP.** The phylogenetic tree (maximal likelihood method) of ISP proteins was computed from the structure-based alignments in Fig. S4 in File S1. The small green circle indicates ancient nitrogen or methylotrophic metabolism [29–32] (Fig. 6B). The dashed green bracket indicates the paralogue proteins belonging to the *bcf* complex. Other brackets indicate proteobacterial subdivisions and mitochondria as in A. Note how the bootstrap values are much lower within the bottom branch containing mitochondrial ISP than in the upper branch containing ISP2. doi:10.1371/journal.pone.0096566.g008

frequency of these characters than  $\alpha$ -proteobacteria (Fig. 6B). However, some  $\alpha$ -proteobacteria show a high frequency of fused proteins and elements of N metabolism (Fig. 6B), namely methylotrophs - encompassing the families of Methylocystaceae, Methylobacteraceae, Beijerinckiaceae and part of Hyphomicrobiaceae, as well as Bradyrhizobiaceae such as *Afipia felis* and *Rhodospseudomonas palustris* BisA53 [77] - several Acetobacteraceae and some Rhodospirillaceae. These organisms also have a wide range of ancestral characters such as type a *COX* operons and ISP2 (Table S2 in File S1 and Fig. 8).

The information just discussed can be integrated with the timeline of bacterial evolution [31], which positions the separation of the  $\beta$ -lineage near the time at which oxygen levels dramatically increased, at least in the photic zone of marine environments and emerged land. The invention of the metabolic pathways of methane, ammonia and nitrite oxidation immediately followed,

allowing autotrophic ways of life which are now retained by a few groups of proteobacteria [30]. These bacteria also possess the largest variety of *COX* operons and molecular forms of their catalytic subunits, as the result of multiple events of operon and gene duplication. Some of these duplications are still evident in extant organisms, as indicated by the doublet of *COX3* proteins in *COX* operon type a-III (Fig. 3A) and the presence of concatenated *COX* operons in some genomes (Table S2 in File S1). Our reconstruction of the molecular evolution of *COX3* proteins and their binding strength for oxygen-modulating phospholipids (Fig. 5) seems to recapitulate a progressive adaptation to increasing levels of  $O_2$ , which had to be gauged in terms of decreasing oxygen affinity to maintain maximal efficiency of the oxidase reactions, with minimal damage by radicals and potential suicidal reactions [47,60,78]. We have also found multiple forms of other terminal oxidases in methylotrophs and Rhodospirillales, in particular for



**Figure 9. Possible progenitors for the bioenergetic evolution of mitochondria.** This diagram is modified from that in Fig. 1B to take into account the deduction that proto-mitochondria probably had two different types of *COX* operons (type a is labelled in dark olive background) and the evidence for multiple ISP forms. ISP2 is represented in a grey box while ISP1 in dark blue. Various steps of differential loss or acquisition via LGT are indicated for the possible pathways of evolution from extant or extinct  $\alpha$ -proteobacteria into proto-mitochondria. By considering the complexities arisen from our data, pathway A in Fig. 1B stemming from *Beijerinckia* would require one loss and one acquisition, while pathway B would theoretically imply two losses and two acquisitions. However, we now exclude that this pathway may have contributed to the evolution of mitochondria (see text). Pathway C, sustained by most results presented here, bypasses the *Beijerinckia* subset with the combined loss of two bioenergetic systems and ISP2. Finally, pathway D would require the combined loss of three bioenergetic systems from organisms such as *Tistrella*, but of two systems plus ISP2 for *R. palustris* BisA53, which has already lost *bo*-type oxidase (Table S1 in File S1). The obvious possibility that yet undiscovered, or extinct bacteria may be among the originators of the proto-mitochondrion is considered, as indicated. Eventual loss of photosynthesis is not shown, but it would apply only to *Methylobacterium*, *R. palustris* and *Roseobacter* among the organisms shown. The grey vertical arrow on the left indicates the possible equivalence of *COX* operon type a with dual function (cytochrome *c* and ubiquinol) oxidases in some Rhodobacterales.

doi:10.1371/journal.pone.0096566.g009

the *bd*-type ubiquinol oxidase (Table S1 in File S1). The additional forms usually correspond to the Cyanide Insensitive Oxidase (CIO) [79], which has lower affinity for oxygen than classical *bd* oxidases [25].

We believe that the large increase in ambient oxygen that occurred during the evolution of primordial proteobacteria [31] was the driving force for the genomic expansion and diversification of oxygen-reacting enzymes. High levels of O<sub>2</sub> also led to the wide availability of nitrate and nitrite that can function as alternative terminal acceptors for electron transfer and bioenergy production [22,32]. This underlines the strong link between oxygen respiration and key elements of N metabolism that we have taken in consideration here. The separation of proto-mitochondria is estimated to have occurred when oxygen levels were still very low in the oceans [10,24], where most primordial life thrived. It is therefore plausible that the distal progenitors of mitochondria were related to organisms that had experimented with a wide variety of oxygen-reacting systems and thus retained great plasticity in their adaptation to micro-oxic or even anoxic environments, a trait that is partially retained in eukaryotes adapted to anaerobic environments [10]. With this conceptual framework in mind, we can now look back to the initial approach of our work (Fig. 1) and consider the most plausible pathways for mitochondrial evolution (Fig. 9).

Following the separation of the  $\beta$ - and  $\gamma$ -proteobacterial lineages, proto-mitochondria may have branched off along one of the pathways illustrated in Fig. 9. Pathways A and B are the same as in Fig. 1B, with the additional complexities that have emerged from the detailed analysis of COX operons and ISP proteins plus possible acquisitions via LGT. Pathway A, stemming from *Beijerinckia* (we now exclude *Micavibrio* for it lacks key elements of N metabolism, cf. Fig. 2), would require one loss (*bd* oxidase) plus one acquisition (COX operon type a-II), while pathways B would theoretically require two losses and two acquisitions of bioenergetic systems. However, our results indicate that mitochondrial evolution is unlikely to have followed pathway B, since the organisms from which it departs do not have key elements of N-metabolism that are present in some eukaryotes (Figs. 2 and 6B) nor a ISP comparable to that of eukaryotes (Figs. 7 and 8). Additional pathway C bypasses the *Beijerinckia* subset with the combined loss of two bioenergetic systems and ISP2, the latter being a facile evolutionary step for only six organisms have retained ISP2 (Figs 7 and 8). This pathway stems from methylotrophic bacteria such as *Methylocystis* and *Methylobacterium*. Indeed, the analysis of three different types of bioenergy-producing systems - cytosolic nitrate assimilation, mitochondria-encoded subunits of cytochrome *c* oxidase and nuclear-encoded ISP subunit of the cytochrome *bc<sub>1</sub>* complex - converges in indicating methylotrophs as the most likely relatives to proto-mitochondria. Moreover, by combining the analysis of nitrate metabolism (Fig. 2) with that of COX (Figs. 3–6) and ISP evolution (Figs. 7, 8 and S4 in File S1), only *Tistrella* [48] and *Rhodospseudomonas palustris* [6] remain among all the bacteria that have been previously proposed as possible ancestors of mitochondria (cf. Figs. 1B and Table S1 in File S1). We have thus considered also pathway D, which would require the combined loss of three bioenergetic systems from those possessed by *Tistrella* (Fig. 9). Finally, *Rhodospseudomonas palustris* BisA53 does not have the *bo*-type oxidase as other organisms of the same genus, but possesses a methanol dehydrogenase close to that of methylotrophs (Table 1). However, it still retains a photosynthetic system, the loss of which would add to the other steps required to resemble proto-mitochondria (Fig. 9). The obvious possibility that yet undiscovered, or extinct bacteria may be among the originators of the proto-mitochondrion is also considered in Fig. 9. Yet, these

unknown organisms would probably have the subsets of bioenergy systems shown in the top part of the diagram.

Taken all our results together, methylotrophic organisms emerge as the closest living models for mitochondrial ancestors. In perspective, our work provides new means for selecting bacterial organisms that are most suitable for experimentally re-evolving proto-mitochondria with mitochondria-depleted eukaryotic cells.

## Methods

To identify genes and their products with others currently present in National Center for Biotechnology Information (NCBI) resources, we have extensively used the program DELTABLAST, Domain Enhanced Lookup time Accelerated BLAST [80], integrated with hydropathy analysis conducted with in house algorithms [72] or the program WHAT (Web-based Hydropathy, Amphipathicity and Topology <http://saier-144-21.ucsd.edu/barwhat.html> [81]). Manually refined alignments of bioenergetic proteins were subjected to phylogenetic analysis with maximum likelihood algorithm and 100 bootstrap resamplings, using the program PhyML 3.0 and evolutionary models selected with Prottest3, as described earlier [21]. The results obtained with this rigorous method essentially matched those obtained with the recent options of DELTABLAST (cf. Fig. 8). The genomes of *Asaia platicody* and *Saccharibacter sp.* (EMBL accession: CBLX010000001/27 and CBLY010000001/09, respectively) were recently reported by Chouaia *et al* [22]. See Supporting Information for additional methods and procedures of gene recognition, operon classification (cf. [82]) and sequence analysis of proteins (cf. [41,52,83]).

## Supporting Information

**File S1 We enclose File S1 with Supporting Information containing a detailed account of the classification of bacterial COX operons (2 pages), 4 additional Figures and 4 additional Tables. Figure S1, Pathways for the bioenergetic evolution of a bacterial not leading to mitochondria.** The diagram shows the additional subsets of bioenergetic systems that are not shown in Fig. 1B, including those of *Asaia* and *Saccharibacter* (Table S1B in File S1). The asterisk\* labels the same subset as in Fig. 1B (main text), but with fewer representative taxa. Underlined organisms are symbionts or pathogens. Each of the six bioenergetic systems presented in Fig. 1 was identified from its catalytic protein subunits and was considered functionally absent when one or more of these subunits were not found in their completeness, as indicated by the profile of their conserved domains (cf. [41]). The functional absence of a given system is represented by an empty square as in Fig. 1B. **Figure S2, Sequence analysis to identify the fusion of COX4 subunit with COX1 proteins. A.** Sequences of recognised or putative COX4 were manually aligned to reference proteins having known 3D structure around the first transmembrane helix (TM1, highlighted in grey): subunit IV of *Thermus caa<sub>3</sub>* oxidase (accession: pdb|2YEV [54]) and subunit IV (COX4<sub>pro\_2</sub> super family [cl06738]) of *Rhodobacter Sphaeroides aa<sub>3</sub>* oxidase (chain D, accession: pdb|1M57 [53]). \*Residues in **bold** have positive scores ( $\geq 0$ ) in the BLOSUM62 substitution matrix [83], those **yellow-highlighted** are identical with either reference protein, while those highlighted in purple are identical to *Janibacter COXIV* (accession: ZP\_00994995) with scores  $\geq 5$  [83]. The total count of identities is also highlighted in yellow (tot) before the description of the protein on the right. It was used to identify other COX4-like proteins such as DUF983 (see Fig. 3A and the section entitled “classification of bacterial COX operons” in File S1). The

minimal count for deeming a protein as “COX4-like” was considered to be 10, but several COX1 proteins exhibited larger numbers of identities. The region of ciliate COX1 showing similarity with COX4 partially overlaps the last transmembrane region (TM12) of aligned COX1, which is well conserved among all available COX1 sequences from ciliates. However, the COX4-like region in bacterial COX1 and that of the pathogenic fungus *Zysoptoria* [55] lies outside the conserved domains of other COX1 proteins. Azospirillum\_bras, *Azospirillum brasilense*; Methylobac\_extor, *Methylobacterium extorquens*. **B** - This panel shows the alignment of COX4 subunits around the second transmembrane helix (TM2), the structure of which is known only for subunit IV of *Thermus caa3* [54] that was used as the reference for aligning bacterial COX4 and mtDNA-encoded proteins. In **bold black** are the residues that are identical in the aligned position of at least two COX4 sequences, or are positive substitutions [83] across at least three aligned COX4 sequences; they are additionally **yellow-highlighted** when identical between at least one bacterial COX4 and one mtDNA-encoded protein (cf. A). In **bold dark blue** are the residues that are positive substitutions between bacterial COX4 and mtDNA-encoded proteins, while those in **bold light blue** are identical or positive substitutions among the aligned mtDNA-encoded proteins. This colour labelling enhances the limited similarity between the sequences shown. **Figure S3, Gene sequence of additional COX operons in diverse bacteria.** The reference gene name for each cluster is indicated on the right of the figure. Symbols identify the same proteins as in Fig. 3A, with the addition of the small gray bar, protein related to nucleotide exchange factor EF-TS. These short proteins were recognised after alignment to the sequence with known 3D structure of Chain A, dimerization domain of Ef-Ts from *Thermus thermophilus* (Accession: pdb|1TFE|A) using a sequence analysis similar to that shown in Fig. S2 in File S1. Hypothetical steps in the evolution of COX operons are indicated. **Figure S4, Structure-based alignment of bacterial and mitochondrial “Rieske” ISP.** The protein sequences of various ISP of the bc<sub>1</sub> complex were aligned following structures available from various sources matching the alignment gaps or insertions with the most refined 3D data [68–71]. The limits of secondary structures (alpha helices, highlighted in green, and beta sheets, highlighted in purple) were deduced from a consensus of the latest coordinates deposited in the NCBI databanks [68–71]. Common insertions and deletions (Indels [72]) between mitochondrial and bacterial sequences are consecutively labelled CIMit1-7 (cf. Fig. 7A). The C terminus of some sequences is truncated at the residue indicated by the numeral before the slash. Key residues for the iron-sulfur cluster, including Y165 influencing its redox potential [71], are in **bold**. Note that *Nitrospira*, *Nitrosomonas*, *Nitrosococcus* and *Methylocystis* are metabolically related by ammonia/methane autothrophy. The organisms follow established phylogenetic distance from top to bottom according to the following taxonomic groups and species. **Cyanobacteria:** Synechocystis (*bef* complex), *Synechocystis* sp. PCC 6803, 192 aa; **Nitrospirales:** *Nitrospira*, *Candidatus Nitrospira defluvi* [73], 183 aa; **ε-proteobacteria:** Epsilon, *Helicobacter pylori*, 167 aa; **Planctomycetes:** *Kuenenia\_2*, *Candidatus Kuenenia stuttgartensis* (in-operon Kuste3096 [66]), 173 aa; *Schlesneria\_2*, *Schlesneria paludicola* DSM 18645 (accession: ZP\_11092182), 189 aa. **γ-proteobacteria:** Nitrosoc, *Nitrosococcus watsonii* C-113, 201 aa; Frateuria, *Frateuria aurantia*, 201 aa; **β-proteobacteria:** Nitrosomonas, *Nitrosomonas europaea* ATCC 19718, 201 aa; Beta, *Neisseria meningitidis* MC58, 193 aa. **α-proteobacteria:** Methylocy\_1 &\_2, *Methylocystis* sp. SC2 [84], \_1 in-operon, 176 aa, \_2 in isolated gene cluster, 209 aa; Methylob\_r, *Methylobacterium radiotolerans* JCM 2831, 189 aa; Nitrobacter,

*Nitrobacter hamburgensis* ISP2, 219 aa; Gluc\_dia, *Gluconacetobacter diazotrophicus* PAL 5 (in isolated gene cluster), 221 aa; Saccharib, *Saccharibacter* sp. (Chouaia *et al.* [22]), 223 aa; Glu\_oxyd, *Gluconobacter oxydans* H24, 218 aa; Beijerinckia, *Beijerinckia indica*, 172 aa; RoseobacterA2, *Roseobacter litoralis petA2* in-operon, 186 aa; Maricaulis\_1, *Maricaulis maris* in-operon, 207 aa; Micavibrio, *Micavibrio aeruginosavorus* [25], 185 aa; Magnetococcus, *Magnetococcus marinus* [76], 178 aa; Rickettsia, *Rickettsia felis*, 177 aa. **Mitochondria: Acanthamoeba**, *Acanthamoeba castellanii*, 235 aa; S\_cerevisiae, *Saccharomyces cerevisiae*, mature 185 aa (3D structure available [85]); Chicken, *Gallus gallus*, mature 192 aa (3D structure available [68]). C-terminal extensions are highlighted in pale blue with some conserved residues in gray. **Table S1, Genomic distribution of bioenergetic systems in α-proteobacteria.** **A.** The genomes of ca. 120 α-proteobacterial organisms were studied from the latest version of the genome NCBI database <http://www.ncbi.nlm.nih.gov/genome/browse/> accessed on 14 March 2014, verifying also the completeness of genomic data (\*). Reconstruction of the various bioenergetic systems (see text) was deduced by combining genomic information with biochemical and microbiological data. The organisms are listed following the left-right sequence in the model of Fig. 1B. Major types of bd oxidases are classified as bd-I or CIO [25,79]. The organisms directly shown in Fig. 1B are yellow highlighted and those proposed to be relatives of mitochondria are shown in italics with pertinent references (including [86,87]). Underlined organisms are symbionts or pathogens. **B.** This table lists the organisms that have been analysed but are not included in the model of Fig. 1B, also because they are in parallel paths of evolution with respect to the mitochondrial subset of bioenergetic systems. The organisms highlighted in pale yellow are shown in Fig. S1 in File S1, while other annotations are the same as in **A**. Complementary information is in Table S2 in File S1. **Table S2, Diverse gene clusters for aa<sub>3</sub>-type oxidase in α-proteobacteria.** The table lists the diverse types of COX operons (Fig. 3A). COX1 proteins recognised as ba3-like\_Oxidase\_I [cd01660] [41] are under the column ba3^ and correspond to class B [26]. Concatenated operons are framed in blue and connected by a thick line. Incomplete (or ‘dead’ [82]) operons, indicated by the asterisk\*, lack one or more of core subunits *ctaC-E* (Fig. 3A). Functional capacity of the oxidase has been deduced also from biochemical studies [88,89]. **Table S3, Phylogenetic distribution of the main characters of COX gene operons.** We constructed a matrix of 11 independent characters (indicated concisely on top of the columns) that could differentiate the gene sequence of COX subunits in the mitochondria of some protists from the gene sequence of bacterial COX operons. The cumulative phenetic analysis indicate that COX operon type a-II of methylotrophs and *Tistrella* (highlighted) share the largest number of characters with COX gene clusters of protist mitochondria (F. Comandatore and C. Bandi, unpublished). **Table S4, Conserved phospholipid binding sites in COX3 proteins.** The alignment in Fig. 4A was enlarged and the residues corresponding to the PL-binding sites and E90 (close to O2 entry in beef COX3 [60]) were considered conserved when producing positive substitutions [83] (bold amino acid symbols in white background). Other substitutions are highlighted in pale brown while identities are identified as **yes**. Organisms are abbreviated as in Fig. 4. (PDF)

## Acknowledgments

We thank Ed Berry (SUNY, Albany, USA), Marta d’Amora, Diego Sona, Alberto Diaspro and Roberto Cingolani (IIT, Italy) for helpful discussion and support.

## Author Contributions

Conceived and designed the experiments: MDE BC FC EC DS CB. Performed the experiments: MDE BC FC EC DS PMJL. Analyzed the data: MDE BC FC DS PMJL CB DD. Contributed reagents/materials/

analysis tools: MDE CB FC CB DD. Wrote the paper: MDE PMJL CB DD. Conceived the work and wrote the bulk of the manuscript: MDE. Inspired various aspects of the work and participated to the writing and refining of the article: CB DD.

## References

- Gray MW (2012) Mitochondrial evolution. *Cold Spring Harb Perspect Biol* 4: a011403.
- Lane N, Martin W (2010) The energetics of genome complexity. *Nature* 467: 929–934.
- Margulis L (1996) Archaeal-eubacterial mergers in the origin of Eukarya: phylogenetic classification of life. *Proc Natl Acad Sci U S A* 93: 1071–1076.
- Andersson SG, Karlberg O, Canbäck B, Kurland CG (2003) On the origin of mitochondria: a genomics perspective. *Philos Trans R Soc Lond B Biol Sci* 358: 165–177.
- Williams KP, Sobral BW, Dickerman AW (2007) A robust species tree for the alphaproteobacteria. *J Bacteriol* 189: 4578–4586.
- Abhishek A, Bavishi A, Bavishi A, Choudhary M (2011) Bacterial genome chimaerism and the origin of mitochondria. *Can J Microbiol* 57: 49–61.
- Georgiades K, Raoult D (2011) The rhizome of *Reclinomonas americana*, *Homo sapiens*, *Pediculus humanus* and *Saccharomyces cerevisiae* mitochondria. *Biol Direct* 6: 55.
- Thiergart T, Landan G, Schenk M, Dagan T, Martin WF (2012) An evolutionary network of genes present in the eukaryote common ancestor: implications for mitochondrial origin. *Genome Biol Evol* 4: 466–485.
- Gribaldo S, Poole AM, Daubin V, Forterre P, Brochier-Armanet C (2010) The origin of eukaryotes and their relationship with the Archaea: are we at a phylogenomic impasse? *Nat Rev Microbiol* 8: 743–752.
- Müller M, Mentel M, van Hellemond JJ, Henze K, Woehle C, et al (2012) Biochemistry and evolution of anaerobic energy metabolism in eukaryotes. *Microbiol Mol Biol Rev* 76: 444–495.
- Searcy DG (2003) Metabolic integration during the evolutionary origin of mitochondria. *Cell Res* 13: 229–238.
- Gabaldón T, Huynen MA (2003) Reconstruction of the proto-mitochondrial metabolism. *Science* 301: 609.
- Brindefalk B, Ettema TJG, Viklund J, Thollesson M, Andersson SGE (2011) A phylometagenomic exploration of oceanic alphaproteobacteria reveals mitochondrial relatives unrelated to the SAR11 clade. *PLOS ONE* 6: e24457.
- Georgiades K, Madoui M-A, Le P Robert C, Raoult D (2011) Phylogenomic analysis of *Odyssella thessalonicensis* fortifies the common origin of Rickettsiales, *Pelagibacter ubique* and *Reclinomonas americana* mitochondrion. *PLOS ONE* 6: e24857.
- Rodríguez-Ezpeleta N, Embley TM (2012) The SAR11 group of alphaproteobacteria is not related to the origin of mitochondria. *PLOS ONE* 7: e30520.
- Esser C, Ahmadijad N, Wiegand C, Rotte C, Sebastiani F, et al (2004) A genome phylogeny for mitochondria among alpha-proteobacteria and a predominantly eubacterial ancestry of yeast nuclear genes. *Mol Biol Evol* 21: 1643–1660.
- Yip C, Harbour ME, Jayawardena K, Fearnley IM, Sazanov LA (2011) Evolution of respiratory complex I: 'supernumerary' subunits are present in the alpha-proteobacterial enzyme. *J Biol Chem* 286: 5023–5033.
- Clements A, Bursac D, Gatsos X, Perry AJ, Covicristov S, et al (2009) The reducible complexity of a mitochondrial molecular machine. *Proc Natl Acad Sci U S A* 106: 15791–15795.
- Davidov Y, Huchon D, Koval SF, Jurkevitch E (2006) A new alpha-proteobacterial clade of *Bdellovibrio*-like predators: implications for the mitochondrial endosymbiotic theory. *Environ Microbiol* 8: 2179–2188.
- Atteia A, Adrait A, Brugière S, Tardif M, van Lis R, et al (2009) A proteomic survey of *Chlamydomonas reinhardtii* mitochondria sheds new light on the metabolic plasticity of the organelle and on the nature of the alpha-proteobacterial mitochondrial ancestor. *Mol Biol Evol* 26: 1533–1548.
- Sassera D, Lo N, Epis S, D'Auria G, Montagna M, et al (2011) Phylogenomic evidence for the presence of a flagellum and *cbb(3)* oxidase in the free-living mitochondrial ancestor. *Mol Biol Evol* 28: 3285–3296.
- Chouaia B, Gaiarsa S, Crotti E, Comandatore F, Degli Esposti M, et al (2014) Acetic acid bacteria genomes reveal functional traits for adaptation to life in insect guts. *Genome Biol Evol*, in press.
- Kim SW, Fushinobu S, Zhou S, Wakagi T, Shoun H (2009) Eukaryotic *nirK* genes encoding copper-containing nitrite reductase: originating from the proto-mitochondrion? *Appl Environ Microbiol* 75: 2652–2658.
- Johnston DT, Wolfe-Simon F, Pearson A, Knoll AH (2009) Anoxygenic photosynthesis modulated Proterozoic oxygen and sustained Earth's middle age. *Proc Natl Acad Sci U S A* 106: 16925–16929.
- Borisov VB, Gennis RB, Hemp J, Verkhovskiy MI (2011) The cytochrome *bd* respiratory oxygen reductases. *Biochim Biophys Acta* 1807: 1398–1413.
- Sousa FL, Alves RJ, Ribeiro MA, Pereira-Leal JB, Teixeira M, et al (2012) The superfamily of heme-copper oxygen reductases: types and evolutionary considerations. *Biochim Biophys Acta* 1817: 629–637.
- Ducluzeau AL, Ouchane S, Nitschke W (2008) The *cbb3* oxidases are an ancient innovation of the domain bacteria. *Mol Biol Evol* 25: 1158–1166.
- McLeod MP, Qin X, Karpathy SE, Gioia J, Highlander SK, et al (2004) Complete genome sequence of *Rickettsia typhi* and comparison with sequences of other rickettsiae. *J Bacteriol* 186: 5842–5855.
- Takaya N (2009) Response to hypoxia, reduction of electron acceptors, and subsequent survival by filamentous fungi. *Biosci Biotechnol Biochem* 73: 1–8.
- Vlaeminck SE, Hay AG, Maignien L, Verstraete W (2011) In quest of the nitrogen oxidizing prokaryotes of the early Earth. *Environ Microbiol* 13: 283–295.
- Battistuzzi FU, Feijao A, Hedges SB (2004) A genomic timescale of prokaryote evolution: insights into the origin of methanogenesis, phototrophy, and the colonization of land. *BMC Evol Biol* 4: 44.
- Simon J, Klotz MG (2013) Diversity and evolution of bioenergetic systems involved in microbial nitrogen compound transformations. *Biochim Biophys Acta* 1827: 114–135.
- Tamas I, Dedysh SN, Liesack W, Stott MB, Alam M, et al (2010) Complete genome sequence of *Beijerinckia indica* subsp. *indica*. *J Bacteriol* 192: 4532–4533.
- Shih PM, Matzke NJ (2013) Primary endosymbiosis events date to the later Proterozoic with cross-calibrated phylogenetic dating of duplicated ATPase proteins. *Proc Natl Acad Sci U S A* 110: 12355–12360.
- Slot JC, Hibbett DS (2007) Horizontal transfer of a nitrate assimilation gene cluster and ecological transitions in fungi: a phylogenetic study. *PLOS ONE* 2: e1097.
- Lin JT, Goldman BS, Stewart V (1994) The *nasFEDCBA* operon for nitrate and nitrite assimilation in *Klebsiella pneumoniae* M5al. *J Bacteriol* 176: 2551–2559.
- Lebrun E, Santini JM, Brugna M, Ducluzeau AL, Ouchane S, et al (2006) The Rieske protein: a case study on the pitfalls of multiple sequence alignments and phylogenetic reconstruction. *Mol Biol Evol* 23: 1180–1191.
- Moreno-Vivián C, Cabello P, Martínez-Luque M, Blasco R, Castillo F (1999) Prokaryotic nitrate reduction: molecular properties and functional distinction among bacterial nitrate reductases. *J Bacteriol* 181: 6573–6584.
- Mohan SB, Schmid M, Jetten M, Cole J (2004) Detection and widespread distribution of the *nrfA* gene encoding nitrite reduction to ammonia, a short circuit in the biological nitrogen cycle that competes with denitrification. *FEMS Microbiol Ecol* 49: 433–443.
- Zhang Y, Rump S, Gladyshev VN (2011) Comparative Genomics and Evolution of Molybdenum Utilization. *Coord Chem Rev* 255: 1206–1217.
- Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, et al (2011) CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res* 39 (Database issue): D225–9.
- Lee SJ, McCormick MS, Lippard SJ, Cho US (2013) Control of substrate access to the active site in methane monooxygenase. *Nature* 494: 380–384.
- Chistoserdova L, Kalyuzhnaya MG, Lidstrom ME (2009) The expanding world of methylotrophic metabolism. *Annu Rev Microbiol* 63: 477–499.
- Stewart JJ, Coyne KJ (2011) Analysis of raphidophyte assimilatory nitrate reductase reveals unique domain architecture incorporating a 2/2 hemoglobin. *Plant Mol Biol* 77: 565–575.
- Lau E, Fisher MC, Stuedler PA, Cavanaugh CM (2013) The methanol dehydrogenase gene, *mxhF*, as a functional and phylogenetic marker for proteobacterial methanotrophs in natural environments. *PLoS ONE* 8: e56993.
- Liu X, Taber HW (1998) Catabolite regulation of the *Bacillus subtilis* *ctaBCDEF* gene cluster. *J Bacteriol* 180: 6154–6163.
- Radzi Noor M, Soulimane T (2012) Bioenergetics at extreme temperature: *Thermus thermophilus* *ba(3)-* and *caa(3)-* type cytochrome *c* oxidases. *Biochim Biophys Acta* 1817: 638–649.
- Burger G, Gray MW, Forget L, Lang BF (2013) Strikingly Bacteria-Like and Gene-Rich Mitochondrial Genomes throughout jakobid protists. *Genome Biol Evol* 5: 418–438.
- Hussain H, Grove J, Griffiths L, Busby S, Cole J (1994) A seven-gene operon essential for formate-dependent nitrite reduction to ammonia by enteric bacteria. *Mol Microbiol* 12: 153–163.
- Refojo PN, Sousa FL, Teixeira M, Pereira MM (2010) The alternative complex III: a different architecture using known building systems. *Biochim Biophys Acta* 1797: 1869–1876.
- Starkenburg SR, Larimer FW, Stein LY, Klotz MG, Chain PS, et al (2008) Complete genome sequence of *Nitrobacter hamburgensis* X14 and comparative genomic analysis of species within the genus *Nitrobacter*. *Appl Environ Microbiol* 74: 2852–2863.
- Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J (2012). The Pfam protein families database. *Nucleic Acids Res* 40 (Database issue): D290–301.
- Svensson-Ek M, Abramson J, Larsson G, Törnroth S, Brzezinski P, et al (2002) The X-ray crystal structures of wild-type and EQ(I-286) mutant cytochrome *c* oxidases from *Rhodospira sphaeroides*. *J Mol Biol* 321: 329–339.
- Lyons JA, Aragão D, Slattery O, Pislakov AV, Soulimane T, et al (2012). Structural insights into electron transfer in *caa3*-type cytochrome oxidase. *Nature* 487: 514–518.

55. Torriani SF, Goodwin SB, Kema GH, Pangilinan JL, McDonald BA (2008) Intraspecific comparison and annotation of two complete mitochondrial genome sequences from the plant pathogenic fungus *Mycosphaerella graminicola*. *Fungal Genet Biol* 45: 628–637.
56. Swart EC, Bracht JR, Magrini V, Minx P, Chen X, et al (2013) The *Oxytricha trifallax* macronuclear genome: a complex eukaryotic genome with 16,000 tiny chromosomes. *PLoS Biol* 11: e1001473.
57. Hane JK, Lowe RG, Solomon PS, Tan KC, Schoch CL, et al (2007) Dothideomycete plant interactions illuminated by genome sequencing and EST analysis of the wheat pathogen *Stagonospora nodorum*. *Plant Cell* 19: 3347–3368.
58. de Graaf RM, van Alen TA, Dutilh BE, Kuiper JW, van Zoggel HJ, et al (2009) The mitochondrial genomes of the ciliates *Euplotes minuta* and *Euplotes crassus*. *BMC Genomics* 10: 514.
59. Tsukihara T, Aoyama H, Yamashita E, Tomizaki T, Yamaguchi H, et al (1996) The whole structure of the 13-subunit oxidized cytochrome c oxidase at 2.8 Å. *Science* 272: 1136–1144.
60. Shinzawa-Itoh K, Aoyama H, Muramoto K, Terada H, Kurauchi T, et al (2007) Structures and physiological roles of 13 integral lipids of bovine heart cytochrome c oxidase. *EMBO J* 26: 1713–1725.
61. Harrenga A, Michel H (1999) The cytochrome c oxidase from *Paracoccus denitrificans* does not change the metal center ligation upon reduction. *J Biol Chem* 274: 33296–33299.
62. Qin L, Hiser C, Mulichak A, Garavito RM, Ferguson-Miller S (2006) Identification of conserved lipid/detergent-binding sites in a high-resolution structure of the membrane protein cytochrome c oxidase. *Proc Natl Acad Sci U S A* 103: 16117–16122.
63. Yanyushin MF, del Rosario MC, Brune DC, Blankenship RE (2005) New class of bacterial membrane oxidoreductases. *Biochemistry* 44: 10037–10045.
64. Brochier-Armanet C, Talla E, Gribaldo S (2009) The multiple evolutionary histories of dioxygen reductases: Implications for the origin and evolution of aerobic respiration. *Mol Biol Evol* 26: 285–297.
65. Thöny-Meyer L (1997) Biogenesis of respiratory cytochromes in bacteria. *Microbiol Mol Biol Rev* 61: 337–376.
66. Kartal B, de Almeida NM, Maalcke WJ, Op den Camp HJ, Jetten MS, et al (2013) How to make a living from anaerobic ammonium oxidation. *FEMS Microbiol Rev* 37: 428–461.
67. Budd A, Devos DP (2012) Evaluating the Evolutionary Origins of Unexpected Character Distributions within the Bacterial Planctomycetes-Verrucomicrobia-Chlamydiae Superphylum. *Front Microbiol* 3: 401.
68. Berry EA, Huang LS, Saechao LK, Pon NG, Valkova-Valchanova M, et al (2004) X-Ray Structure of *Rhodobacter Capsulatus* Cytochrome bc(1): Comparison with its Mitochondrial and Chloroplast Counterparts. *Photosynth Res* 81: 251–275.
69. Esser L, Elberry M, Zhou F, Yu CA, Yu L, et al (2008) Inhibitor-complexed structures of the cytochrome bc1 from the photosynthetic bacterium *Rhodobacter sphaeroides*. *J Biol Chem* 283: 2846–2857.
70. Zhang Z, Huang L, Shulmeister VM, Chi YI, Kim KK, et al (1998) Electron transfer by domain movement in cytochrome bc1. *Nature* 392: 677–684.
71. Kolling DJ, Brunzelle JS, Lhee S, Crofts AR, Nair SK (2007) Atomic resolution structures of rieske iron-sulfur protein: role of hydrogen bonds in tuning the redox potential of iron-sulfur clusters. *Structure* 15: 29–38.
72. Degli Esposti M, De Vries S, Crimi M, Ghelli A, Patarnello T, et al (1993). Mitochondrial cytochrome b: evolution and structure of the protein. *Biochim Biophys Acta* 1143: 243–271.
73. Valas RE, Bourne PE (2009) Structural analysis of polarizing indels: an emerging consensus on the root of the tree of life. *Biol Direct* 4: 30.
74. Ouchane S, Nitschke W, Bianco P, Vermeglio A, Astier C (2005) Multiple Rieske genes in prokaryotes: exchangeable Rieske subunits in the cytochrome bc-complex of *Rubrivivax gelatinosus*. *Mol Microbiol* 57: 261–275.
75. Lücker S, Wagner M, Maixner F, Pelletier E, Koch H, et al (2010) A *Nitrospira* metagenome illuminates the physiology and evolution of globally important nitrite-oxidizing bacteria. *Proc Natl Acad Sci U S A* 107: 13479–13484.
76. Schübbe S, Williams TJ, Xie G, Kiss HE, Brettin TS, et al (2009) Complete genome sequence of the chemolithoautotrophic marine magnetotactic coccus strain MC-1. *Appl Environ Microbiol* 75: 4835–4852.
77. Simmons SS, Isokpehi RD, Brown SD, McAllister DL, Hall CC, et al (2011) Functional Annotation Analytics of *Rhodospseudomonas palustris* Genomes. *Bioinform Biol Insights* 5: 115–129.
78. Bratton MR, Pressler MA, Hosler JP (1999) Suicide inactivation of cytochrome c oxidase: catalytic turnover in the absence of subunit III alters the active site. *Biochemistry* 38: 16236–16245.
79. Cunningham L, Pitt M, Williams HD (1997) The cioAB genes from *Pseudomonas aeruginosa* code for a novel cyanide-insensitive terminal oxidase related to the cytochrome bd quinol oxidases. *Mol Microbiol* 24: 579–591.
80. Boratyn GM, Schäffer AA, Agarwala R, Altschul SF, Lipman DJ, et al (2012). Domain enhanced lookup time accelerated BLAST. *Biol Direct* 7: 12.
81. Cuff JA, Clamp ME, Siddiqui AS, Finlay M, Barton GJ (1998) *Bioinformatics* 14: 892–893.
82. Price MN, Arkin AP, Alm EJ (2006) The life-cycle of operons. *PLoS Genet* 2: e96.
83. Hung CL, Lee C, Lin CY, Chang CH, Chung YC, et al (2010) Feature amplified voting algorithm for functional analysis of protein superfamily. *BMC Genomics* 11 (Suppl 3): S14.
84. Dam B, Dam S, Kube M, Reinhardt R, Liesack W (2012) Complete genome sequence of *Methylocystis* sp. strain SC2, an aerobic methanotroph with high-affinity methane oxidation potential. *J. Bacteriol* 194: 6008–6009.
85. Lange C, Hunte C (2002) Crystal structure of the yeast cytochrome bc1 complex with its bound substrate cytochrome c. *Proc Natl Acad Sci U S A* 99: 2800–2805.
86. Yang D, Oyaizu Y, Oyaizu H, Olsen GJ, Woese CR (1985) Mitochondrial origins. *Proc Natl Acad Sci U S A* 82: 4443–4447.
87. Fitzpatrick DA, Creevey CJ, McInerney JO (2006) Genome phylogenies indicate a meaningful alpha-proteobacterial phylogeny and support a grouping of the mitochondria with the Rickettsiales. *Mol Biol Evol* 23: 74–85.
88. Sakurai K, Arai H, Ishii M, Igarashi Y (2011) Transcriptome response to different carbon sources in *Acetobacter acetii*. *Microbiology* 157 (Pt 3): 899–910.
89. Gómez-Manzo S, Chavez-Pacheco JL, Contreras-Zentella M, Sosa-Torres ME, Arreguín-Espinosa R, et al (2010) Molecular and catalytic properties of the aldehyde dehydrogenase of *Gluconacetobacter diazotrophicus*, a quinoheme protein containing pyrroloquinoline quinone, cytochrome b, and cytochrome c. *J Bacteriol* 192: 5718–5724.